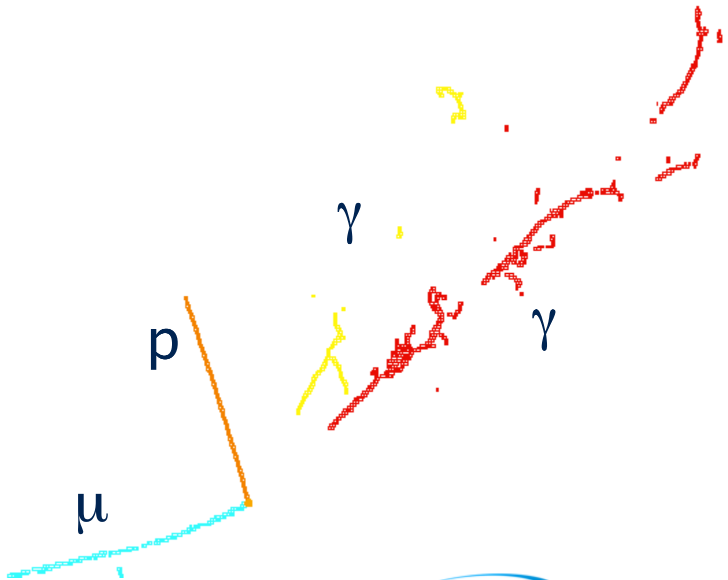


Pandora pattern recognition for LArTPCs

L. Escudero
for the Pandora Team
& the MicroBooNE and DUNE collaborations

PONDD Workshop
6th of December 2018

Thanks to John Marshall, Andy Blake and Steve Green for contributing (voluntarily or involuntarily) to these slides

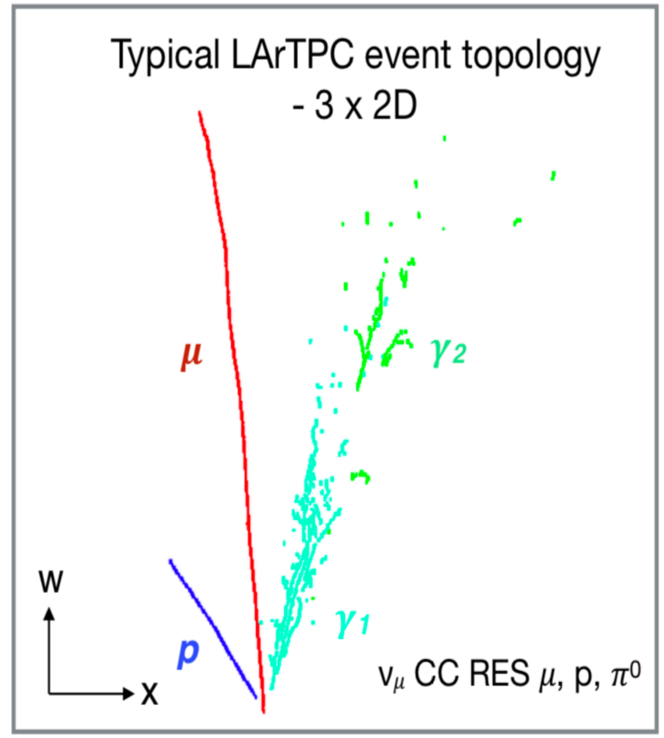
Pandora Pattern Recognition

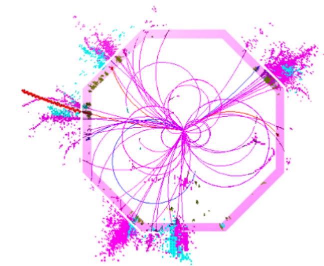
A single clustering approach is unlikely to work for complex topologies (mix of track and shower-like clusters)

Instead, the Pandora project is a novel method of pattern recognition, which tackles this project from its beginning in ILC and LHC using an advanced multi-algorithm approach:

- **Build up events gradually**
- **Each step is incremental - aim not to make mistakes (hard to undo)**
- **Deploy more sophisticated algorithms as picture develops**
- **Build physics and detector knowledge into algorithms**

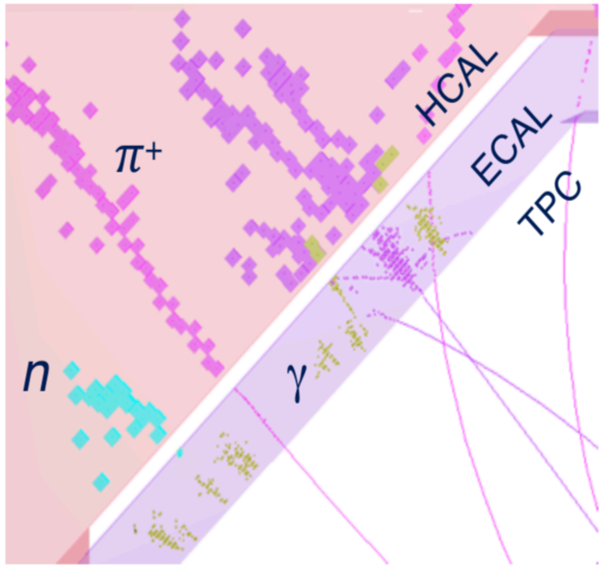
- **Possible thanks to the Pandora Software Development Kit for Pattern Recognition (Eur. Phys. J. C 2015, 75: 439) for all use cases**

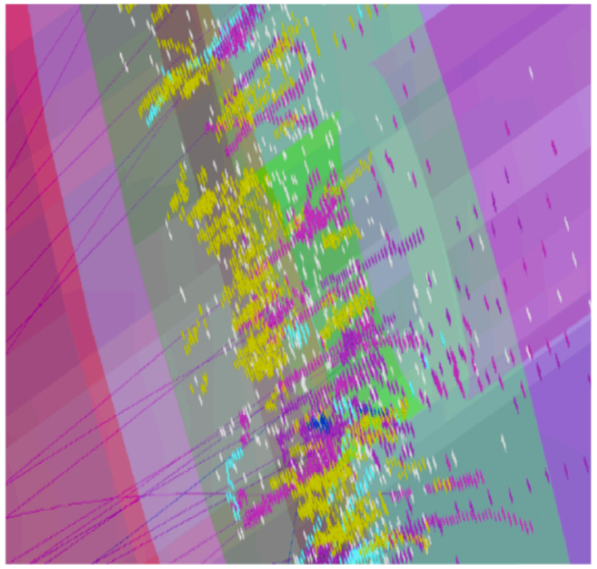




John Marshall

Mark Thomson Andy Blake





Typical ILC event topologies - 3D
NIMA.2009.09.009 NIMA.2012.10.038

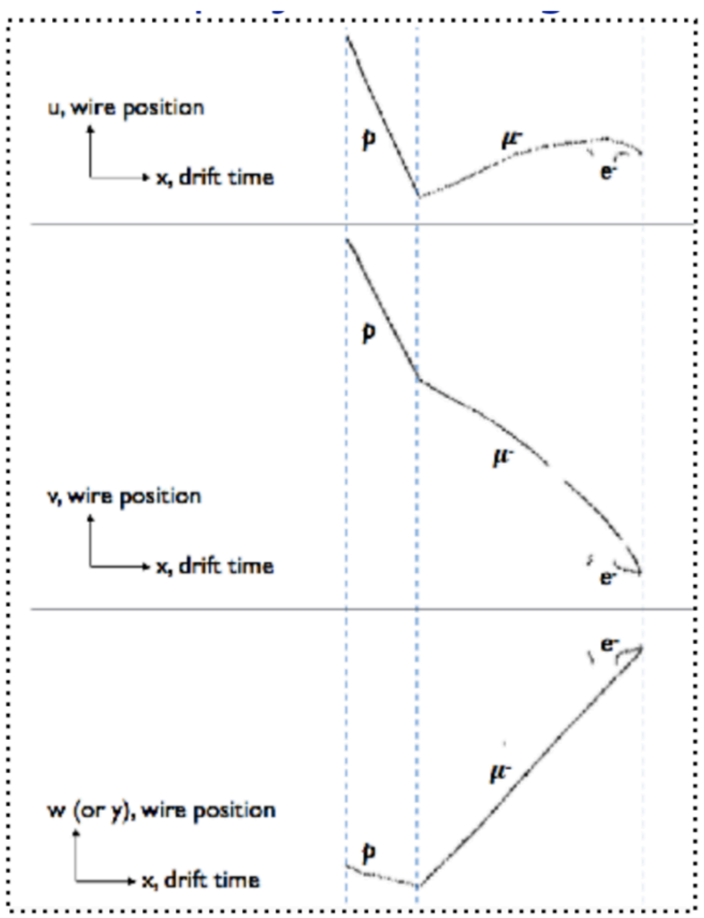
Typical showers in CMS HGCAL - 3D
LHCC-P-008

Pattern Recognition Stage

Low level Reconstruction
Noise Filtering
Signal Processing
Hit Reconstruction

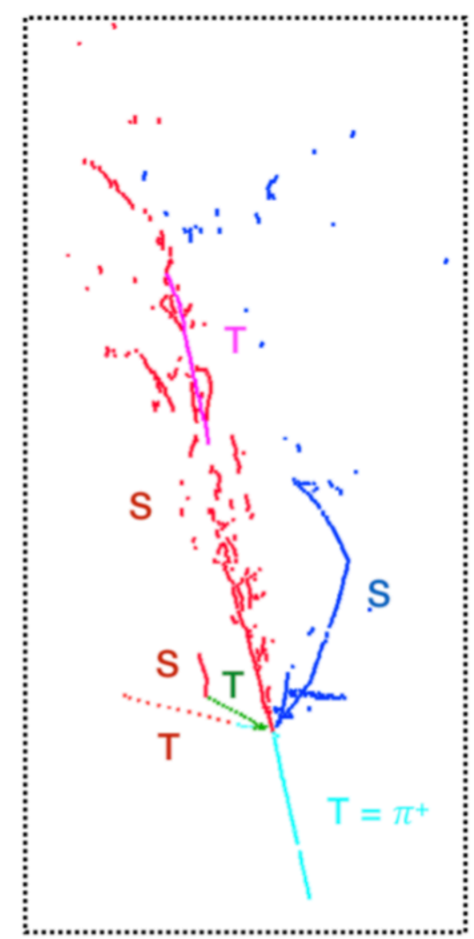
Reconstruction path to Physics

High level Reconstruction
Track Fitting
Calorimetry
Particle ID



This is us!

Pattern Recognition
Clustering
2D -> 3D
Particle Hierarchy



Our **input**: collection of 2D hits

Our **output**: hierarchy of reconstructed particles

Pattern Recognition Stage

Input: 2D hits

2D clustering

3D vertexing

2D/3D matching

Particle refinement

Particle hierarchies

Output: 3D particles

This is us!

Pattern Recognition
Clustering
2D -> 3D
Particle Hierarchy

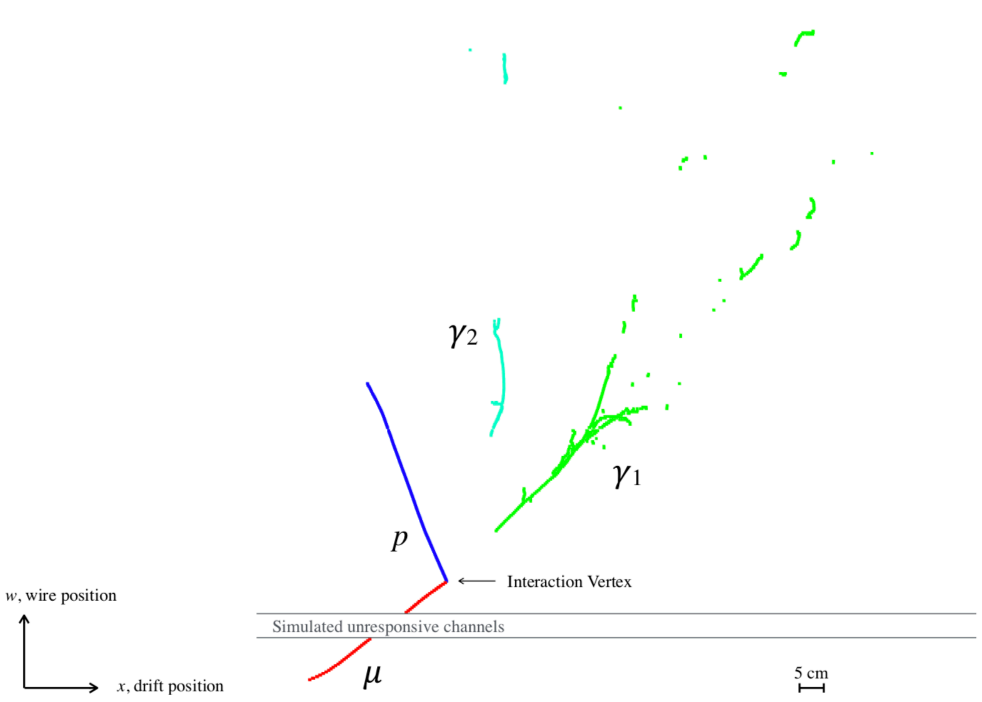
Building up events gradually, with chains of small algorithms, harnessing a number of powerful capabilities:

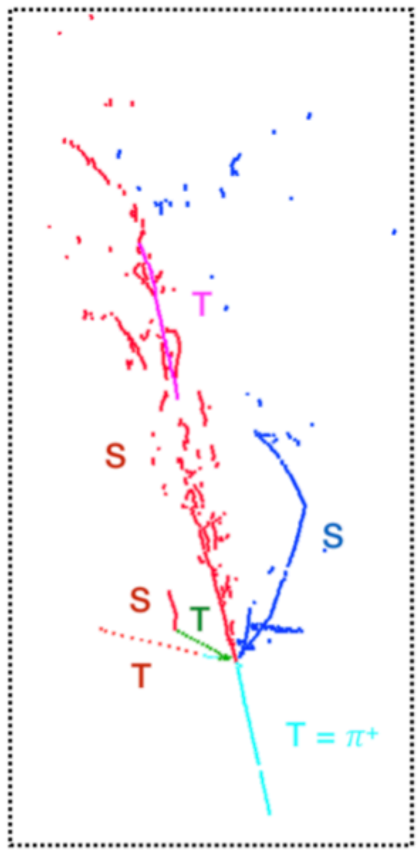
- 120+ algorithms and tools
- Use of multiple parallel hypothesis
- Iterative reconstruction techniques
- Allowing incorporation of ML/DL methods to make algorithm decisions

Pandora for LAr TPCs

- At present Pandora's main development is for pattern recognition in Liquid Argon Time Projection Chambers (LAr TPCs) technology
- Pandora LArTPC algorithms are designed to be reusable for different single-phase LArTPCs, with some retuning expected (e.g. due to different neutrino energies)
- Focus in recent years mainly in MicroBooNE and protoDUNE, but expanded to also SBND, ICARUS, DUNE FD

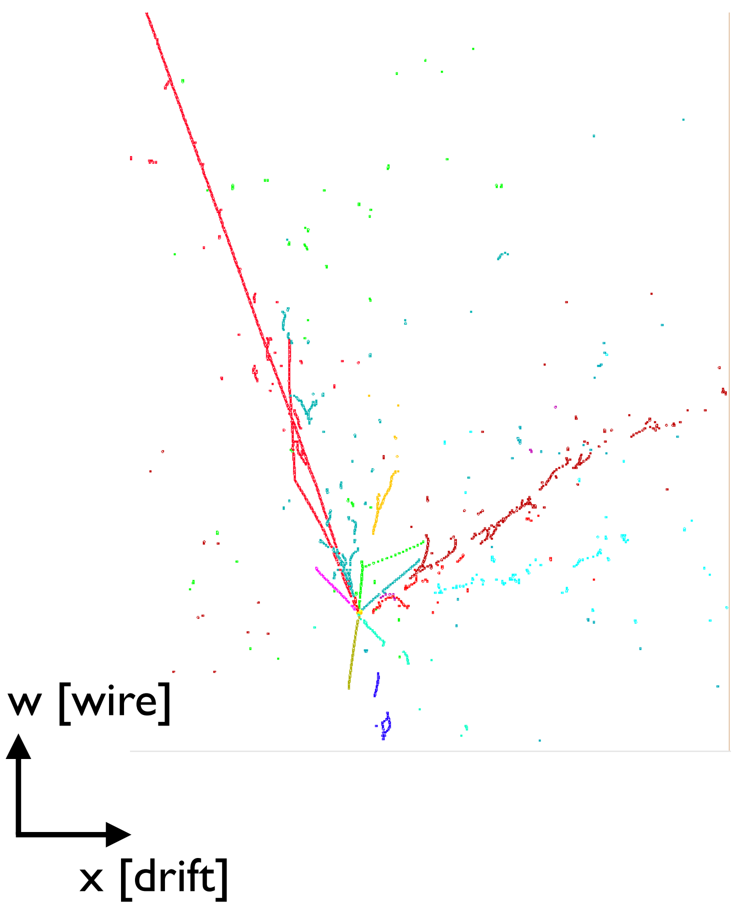
MicroBooNE simulation





Simulated π^+ Pandora Reconstruction at ProtoDUNE-SP

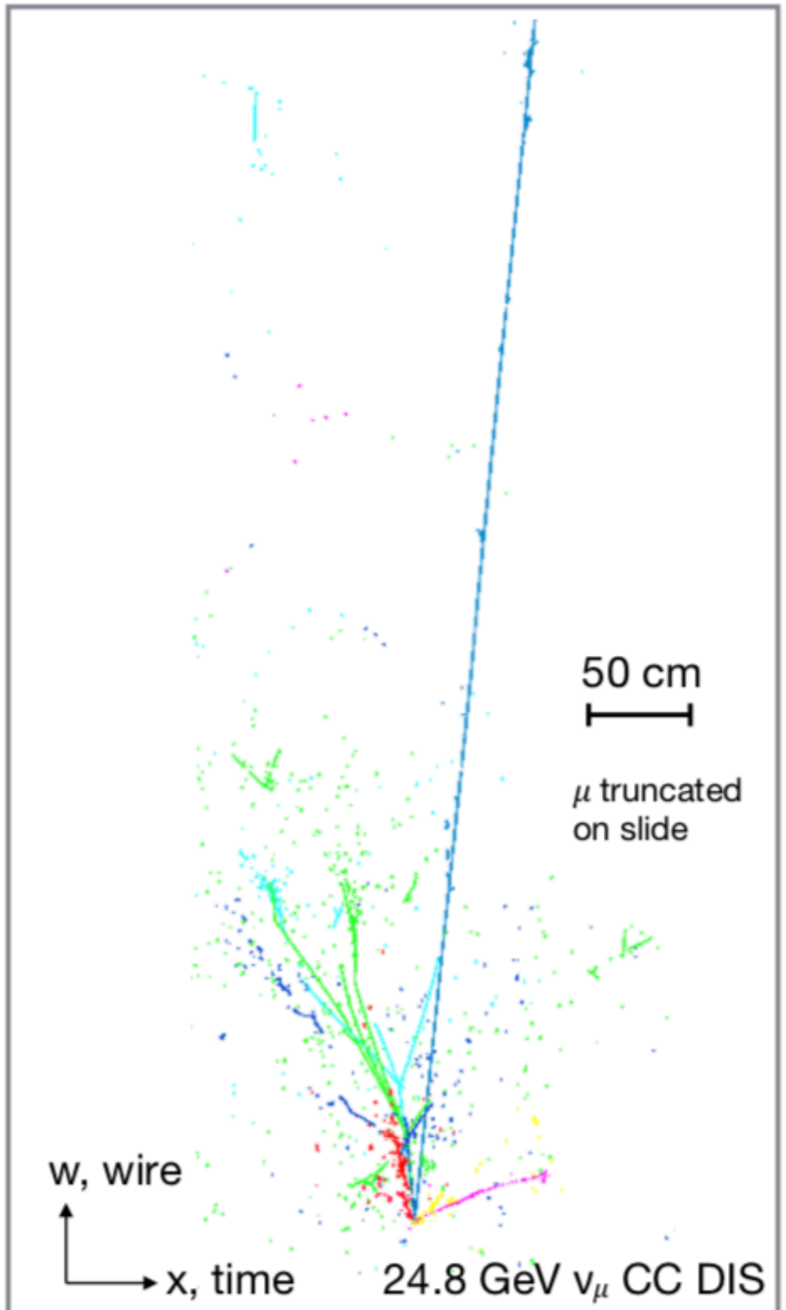
DUNE FD simulation

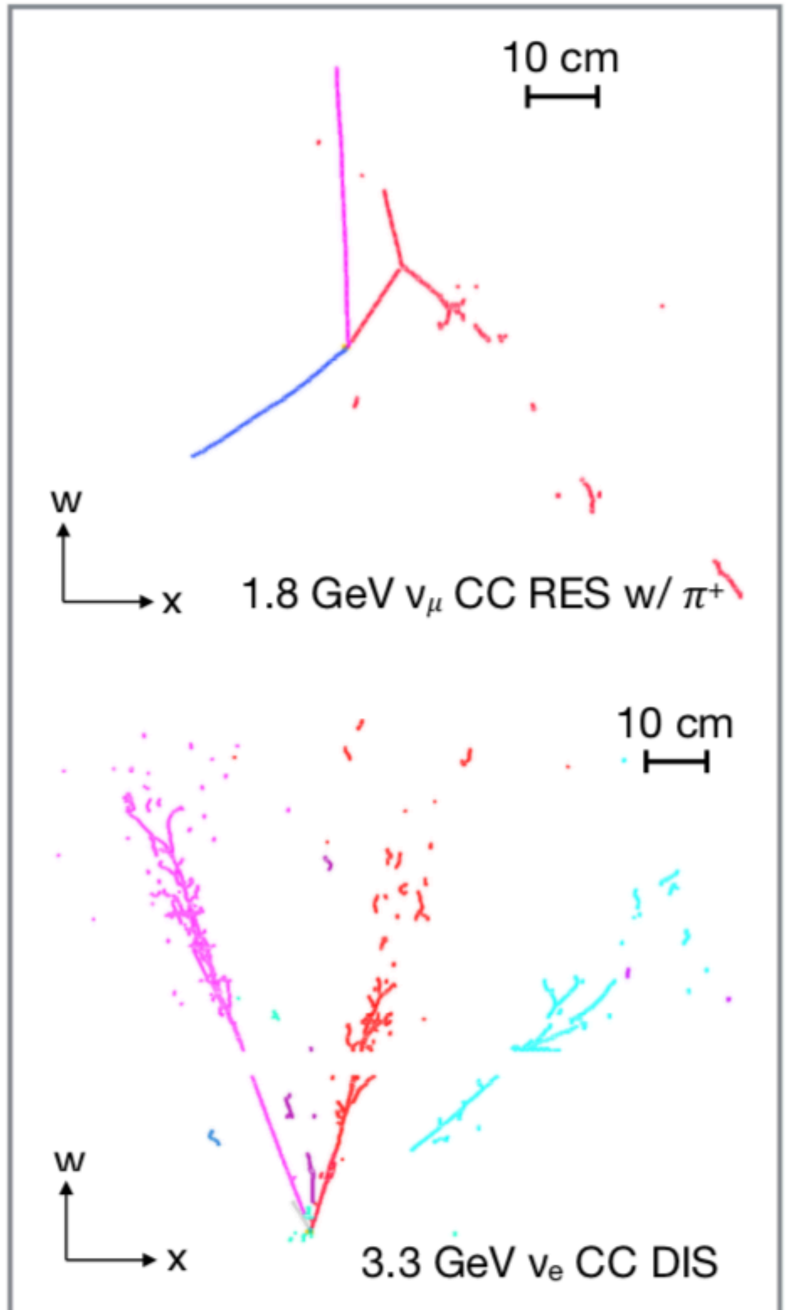


Pandora Pattern Recognition Aim

Our aim is to provide automated pattern recognition for general usage (any particle, any topology)

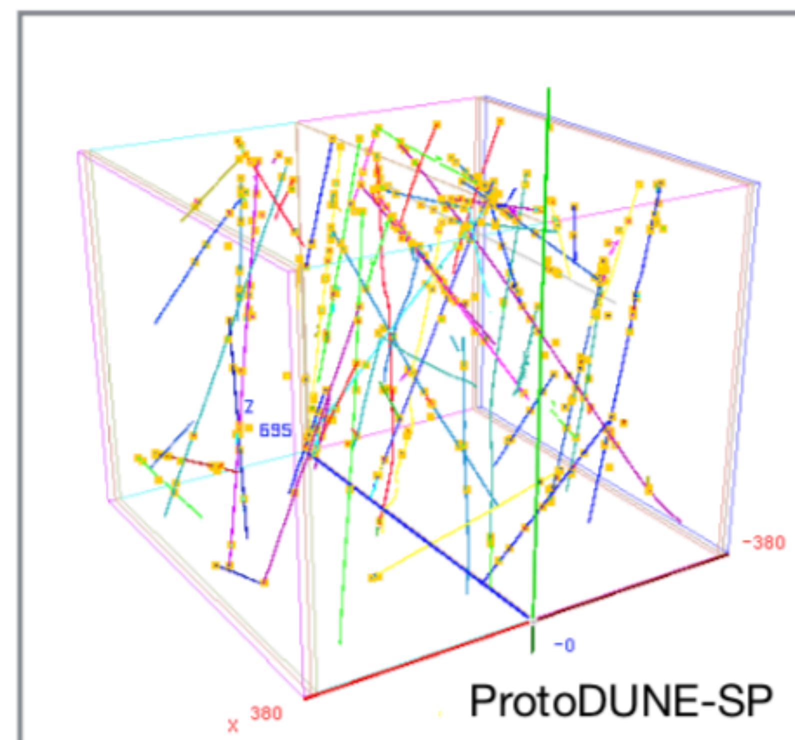
Reconstructing these images (i.e. transforming them into analysis-level physics quantities) is a significant (and exciting!) challenge





LArTPC offer high quality images of complex, diverse topologies.

In addition, they have lengthy drift times, i.e. long exposures (up to few ms) which means a significant cosmic-ray background for surface detectors.



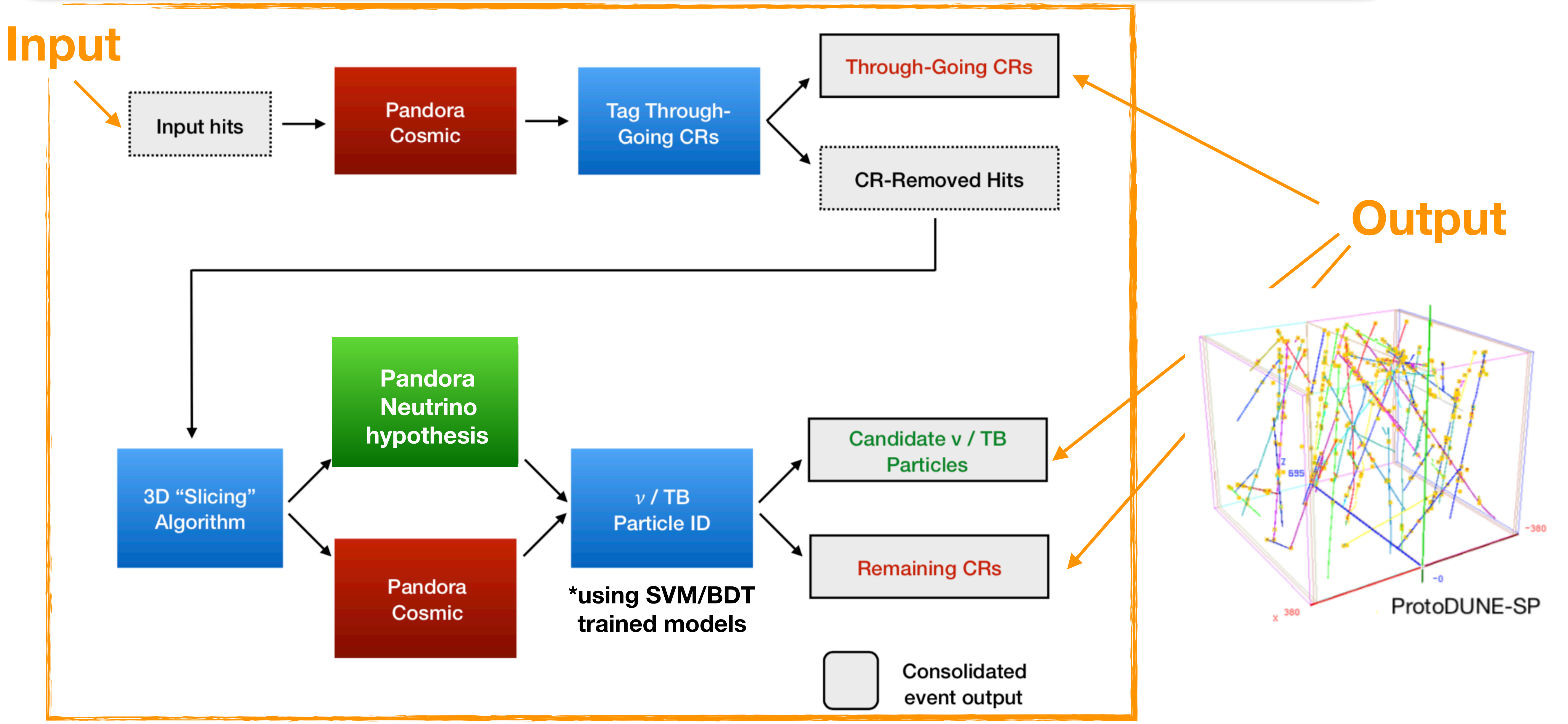
Pandora Consolidated Output

Two different chains with tens of algorithms each:

- **Cosmic ray reconstruction**
- **Neutrino interaction/test beam reconstruction**

Andy Smith

Harnessing the chains of algorithms in an intelligent manner to provide a consolidated output:



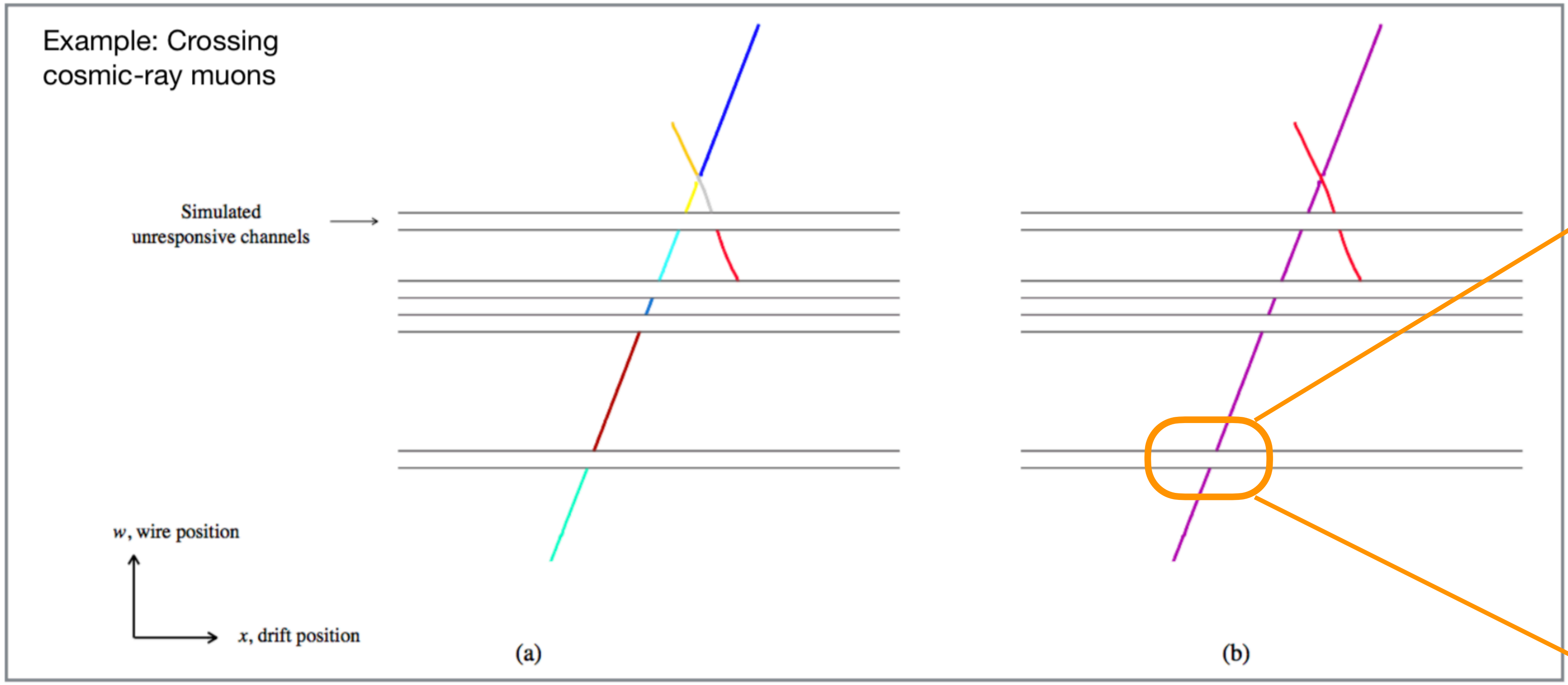
Pandora LAr TPC Pattern Recognition

A little bit deeper into some of the steps in the algorithm chains…

2D track reconstruction

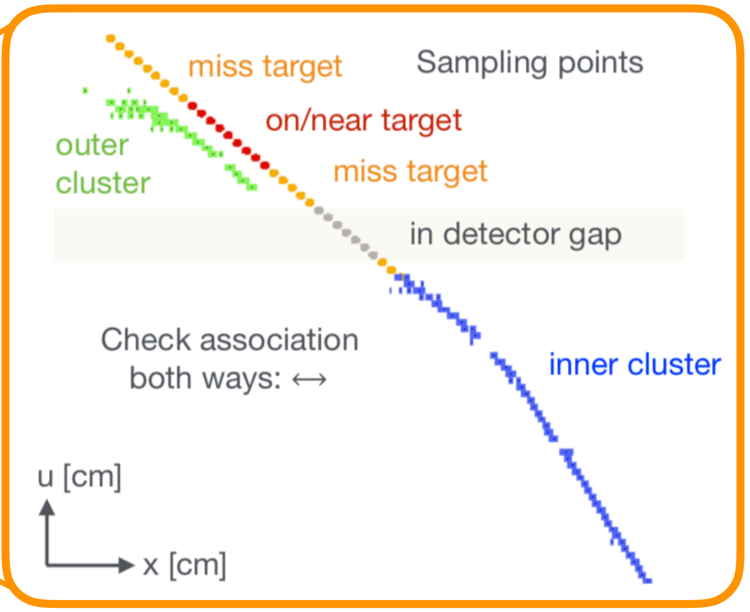
a) We start by producing a list of 2D clusters (per plane) that represent continuos, unambiguous lines of hits, starting/stopping at each branch or ambiguity.

b) Then series of cluster-merging and cluster-splitting algorithms evaluate the list of 2D clusters and change them based on topological information, carefully aiming at safe merges, improving completeness without compromising purity.



Example of cluster-merging algorithm "in action"

E.g. CrossGapsAssociation



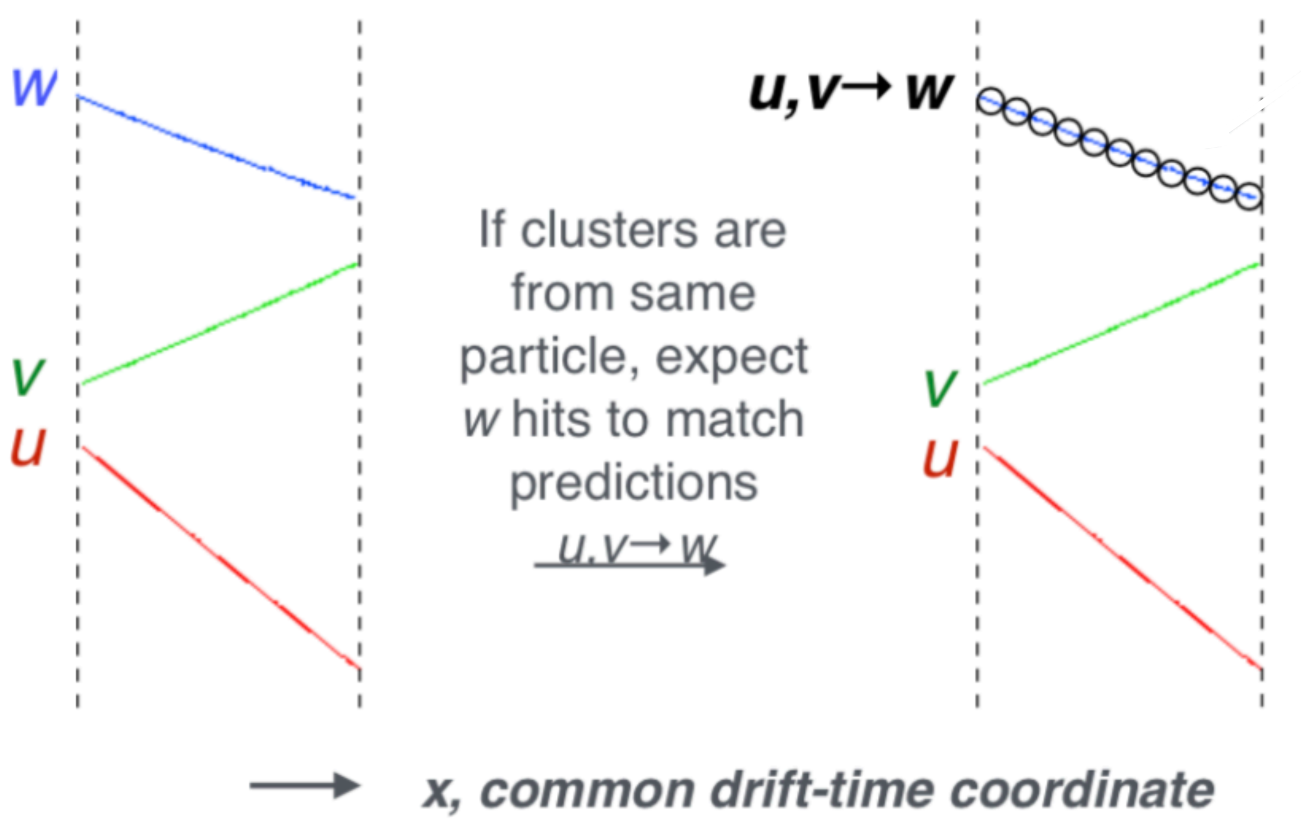
More details on algorithm description in the Pandora MicroBooNE paper [(Eur. Phys. J. C 78, p82 2018)]

Pandora LAr TPC Pattern Recognition

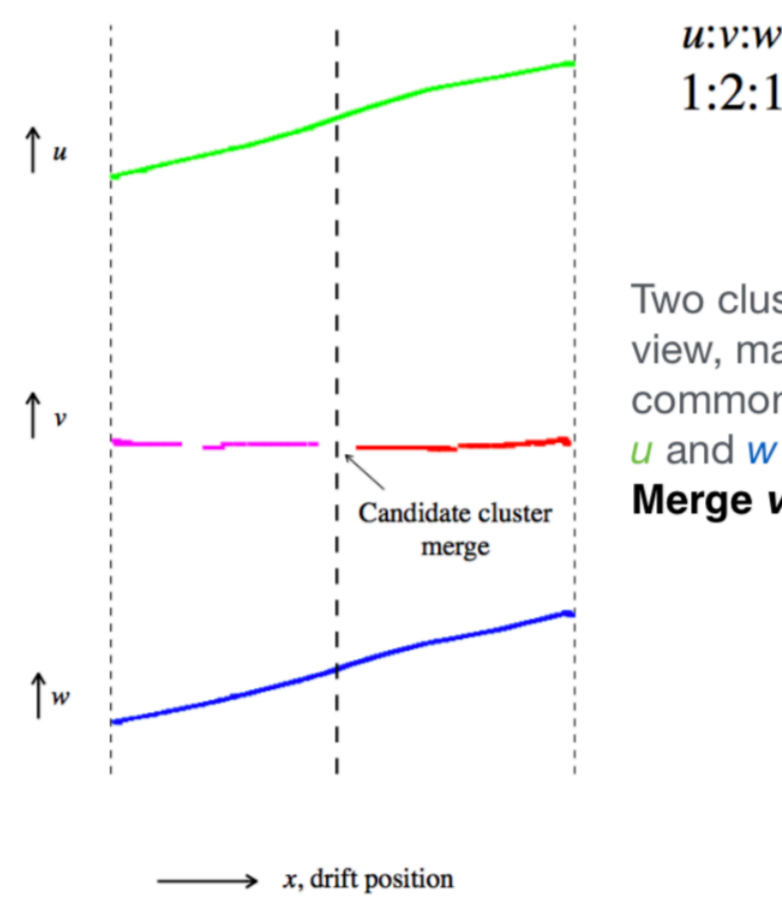
A little bit deeper into some of the steps in the algorithm chains…

3D track reconstruction

2D clusters in the three planes are compared to find those representing same particle, exploiting the common drift-time coordinate and our understanding of wire plane geometry. Results are stored for each 3x2D combination of clusters in a rank-three tensor.



Cluster-matching ambiguities are identified by tools and used to “diagonalise” the tensor **Tools modify 2D clusters as appropriate and then run again from the beginning on the updated tensor. Example tool:**



Two clusters in *v* view, matched to common clusters in *u* and *w* views. **Merge *v* clusters.**

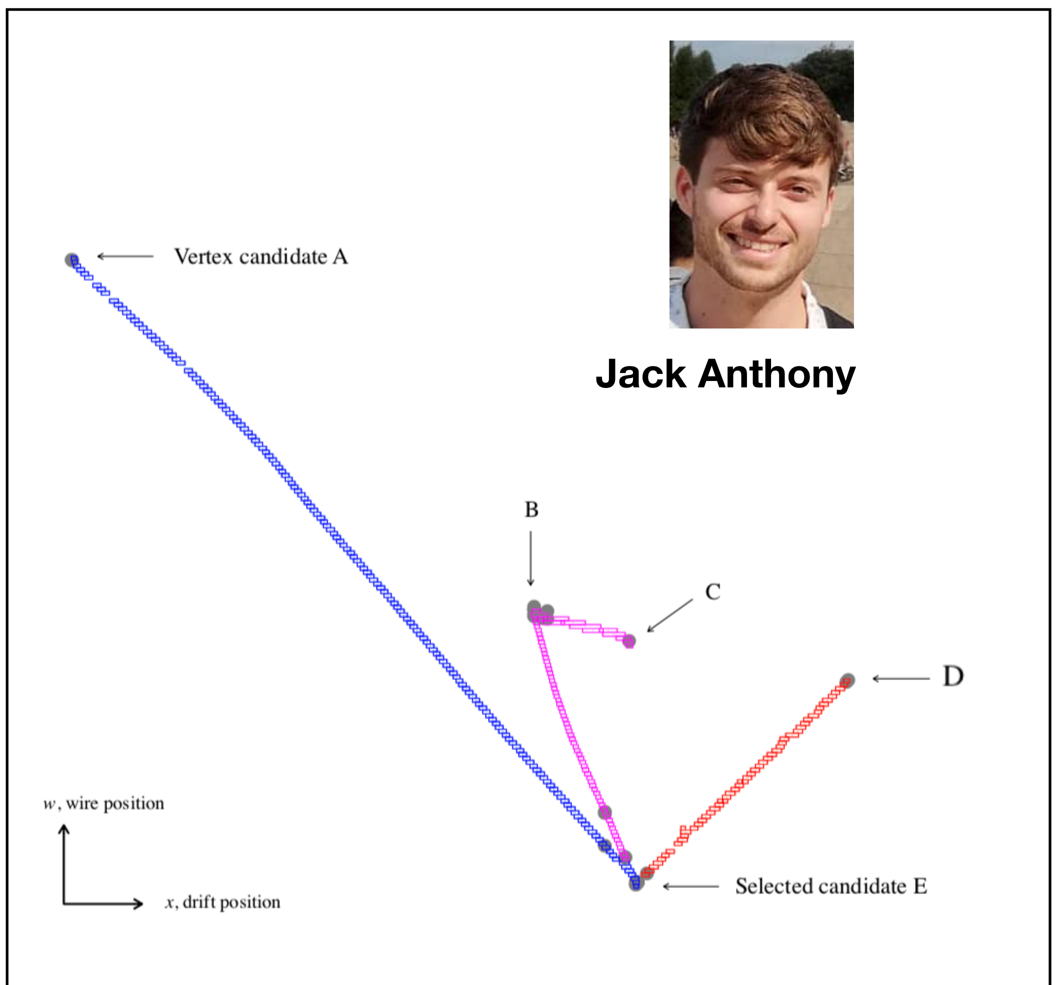
Undershoot Tracks Tool

Pandora LAr TPC Pattern Recognition

A little bit deeper into some of the steps in the algorithm chains…

Neutrino Interaction Vertex

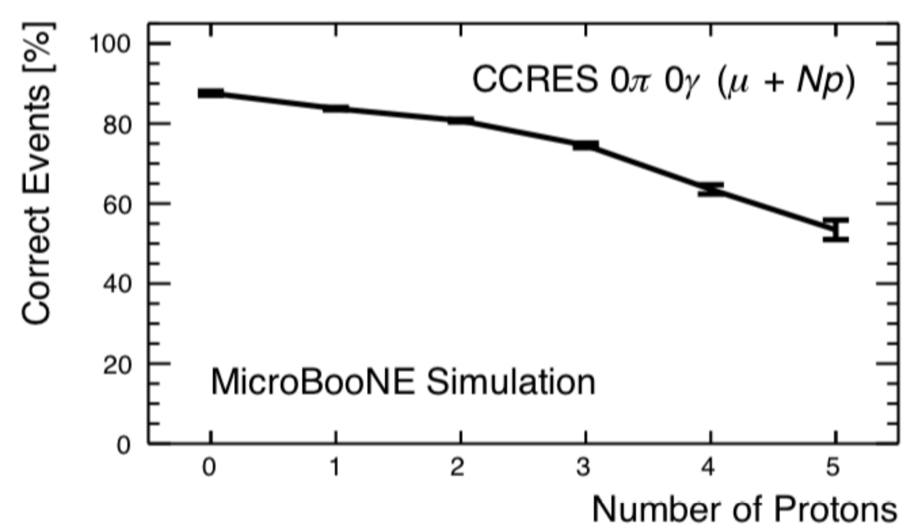
A key algorithm is the one to select the most appropriate 3D vertex position from a list of candidate vertices. Used first a simple score, then a more sophisticated one (with topological and charge asymmetry information) for each candidate. Now a multivariate approach (SVMs) is used for MicroBooNE. This is an example of how powerful the multi-algorithm approach is, by breaking down pattern recognition into small problems, even allowing to use Machine and Deep Learning methods to solve some of them!



Downstream usage:

- **Split 2D clusters at projected vertex position, and use vertex to protect primary particles**

This allows us to have good performance in interactions with many final state particles…



(more later!)

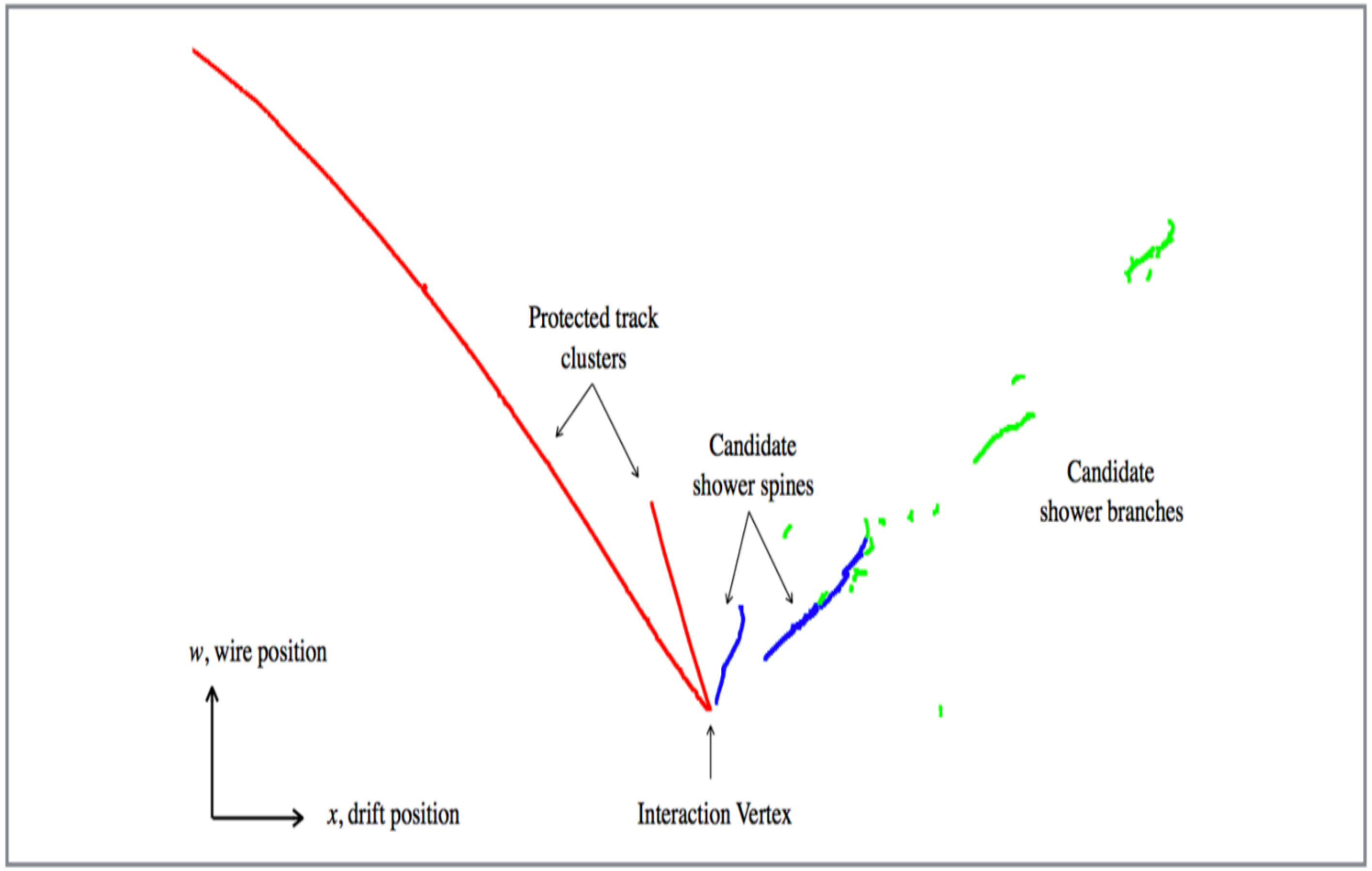
Pandora LAr TPC Pattern Recognition

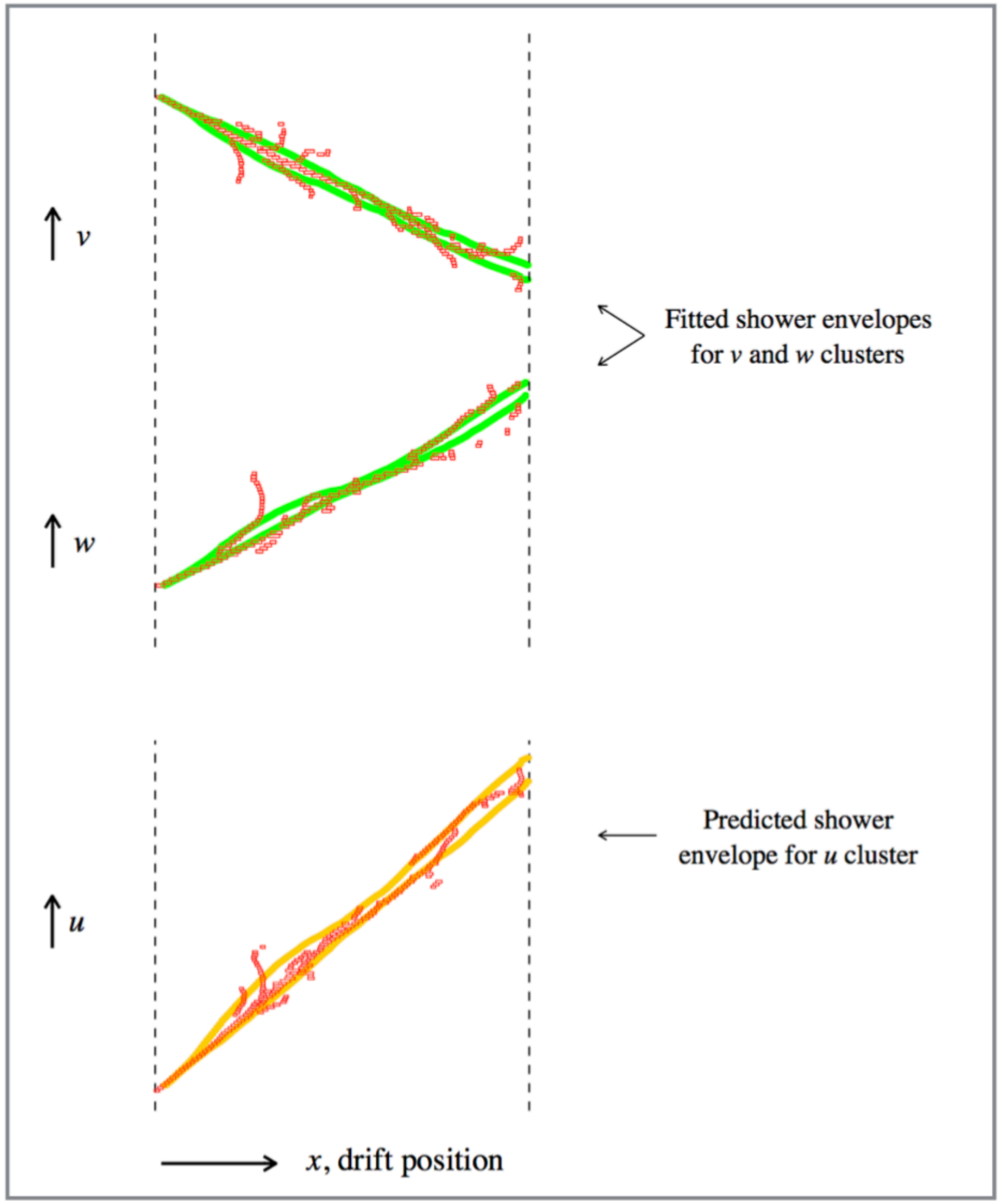
A little bit deeper into some of the steps in the algorithm chains…

Shower Reconstruction

2D - Characterise 2D clusters as track/shower like using topological information, to identify shower spines and allow to grow branches (nearby shower-like clusters), whereas prevent doing so around track-like clusters.

3D - Reuse ideas from the track tensor, using envelopes and their projection to match 2D shower clusters





Pandora LAr TPC Pattern Recognition

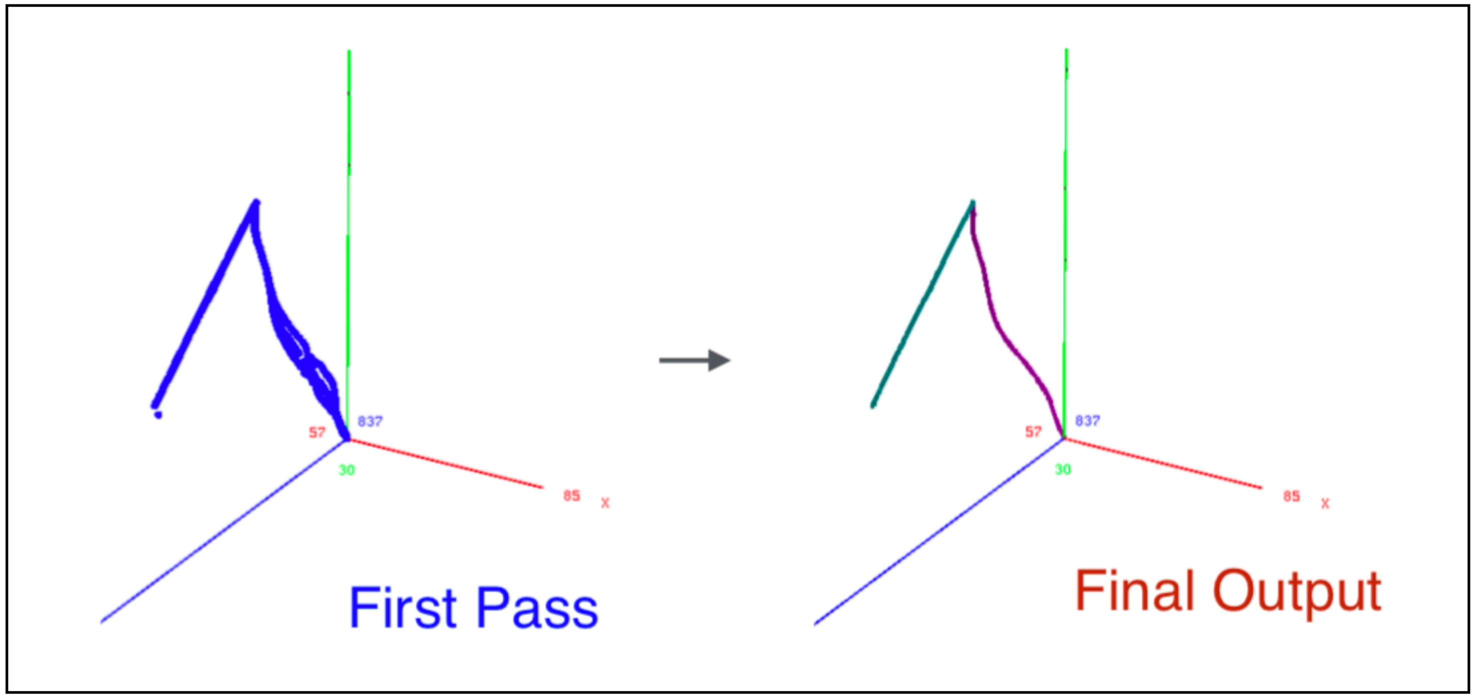
A little bit deeper into some of the steps in the algorithm chains…

3D hit/cluster reconstruction

For each 2D hit, sample clusters in other views at the same X, find u_{in}, v_{in}, w_{in} and use analytic expression to find the most consistent 3D space point by minimising:

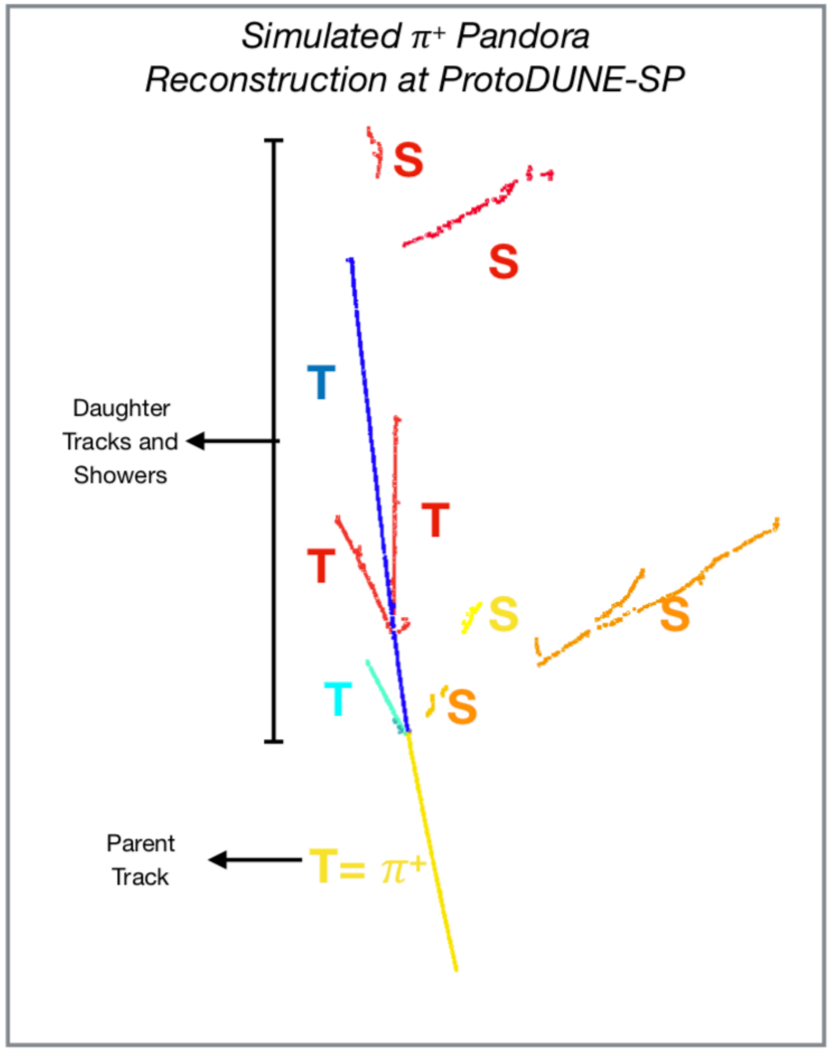
$$\chi^2 = (u_{out} - u_{in})^2 / \sigma_u^2 + (v_{out} - v_{in})^2 / \sigma_v^2 + (w_{out} - w_{in})^2 / \sigma_w^2$$

Iteratively, using fit to current 3D hits, produce smooth trajectory:



and 3D particle hierarchy

Finally, walking backwards from interaction vertex, use 3D clusters to organise particles into hierarchies (building parent-daughter links)

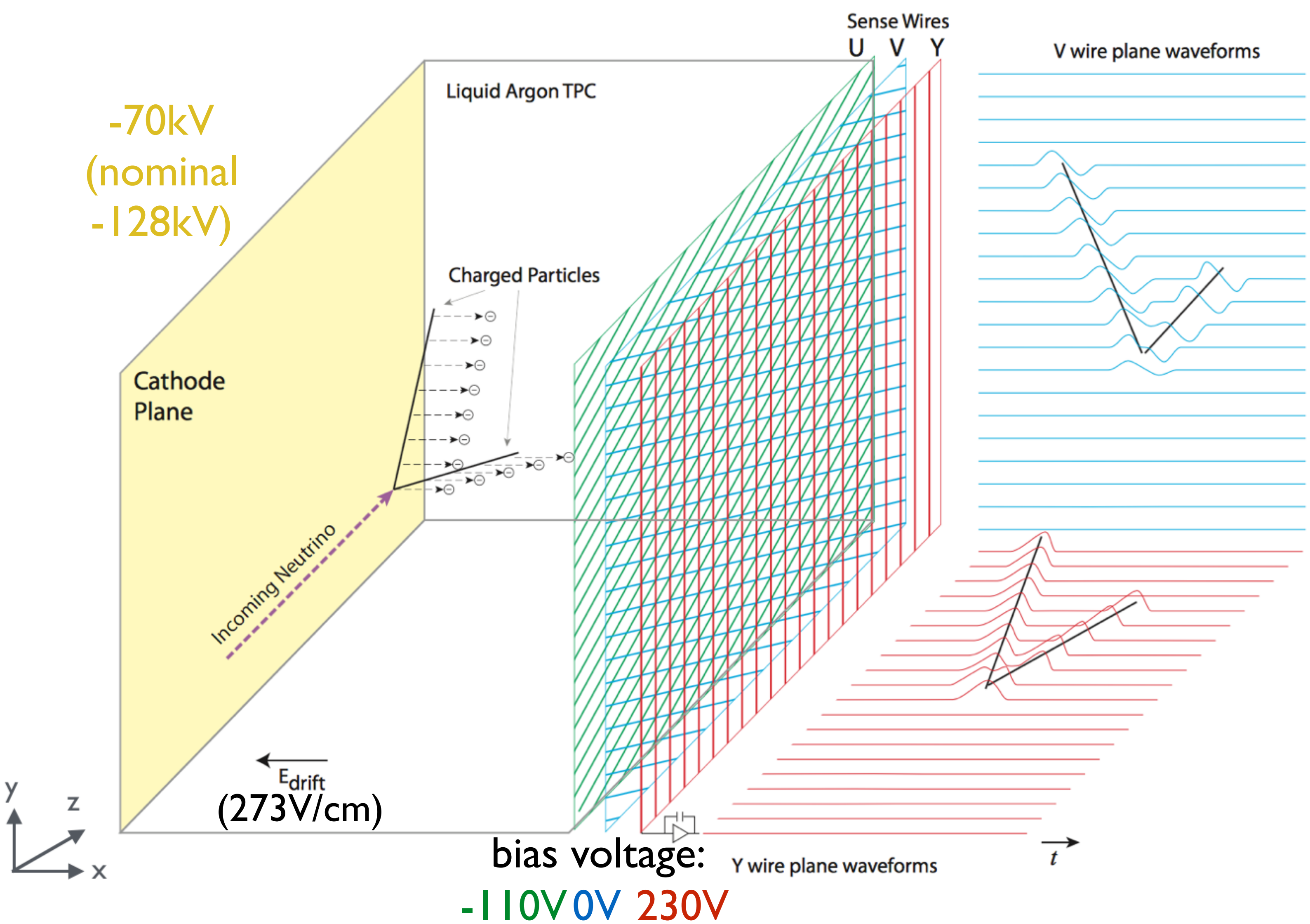


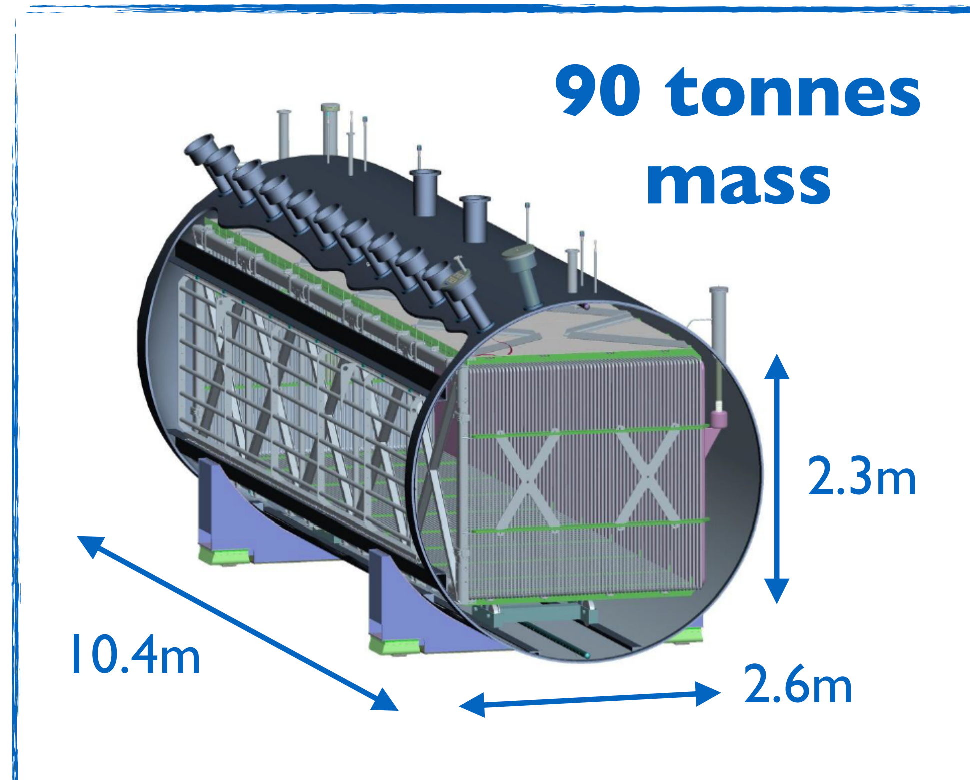
Can't do justice in a few slides, please find more in the Pandora MicroBooNE paper (Eur. Phys. J. C 78, p82 2018)

The MicroBooNE experiment

8256 (3 mm pitch) sense wires in 3 planes:

- The first two planes (U and V) are induction planes with wire orientation +/-60⁰.
- The third plane (Y=W) is the collection plane with wires oriented vertically





Surface LAr TPC detector

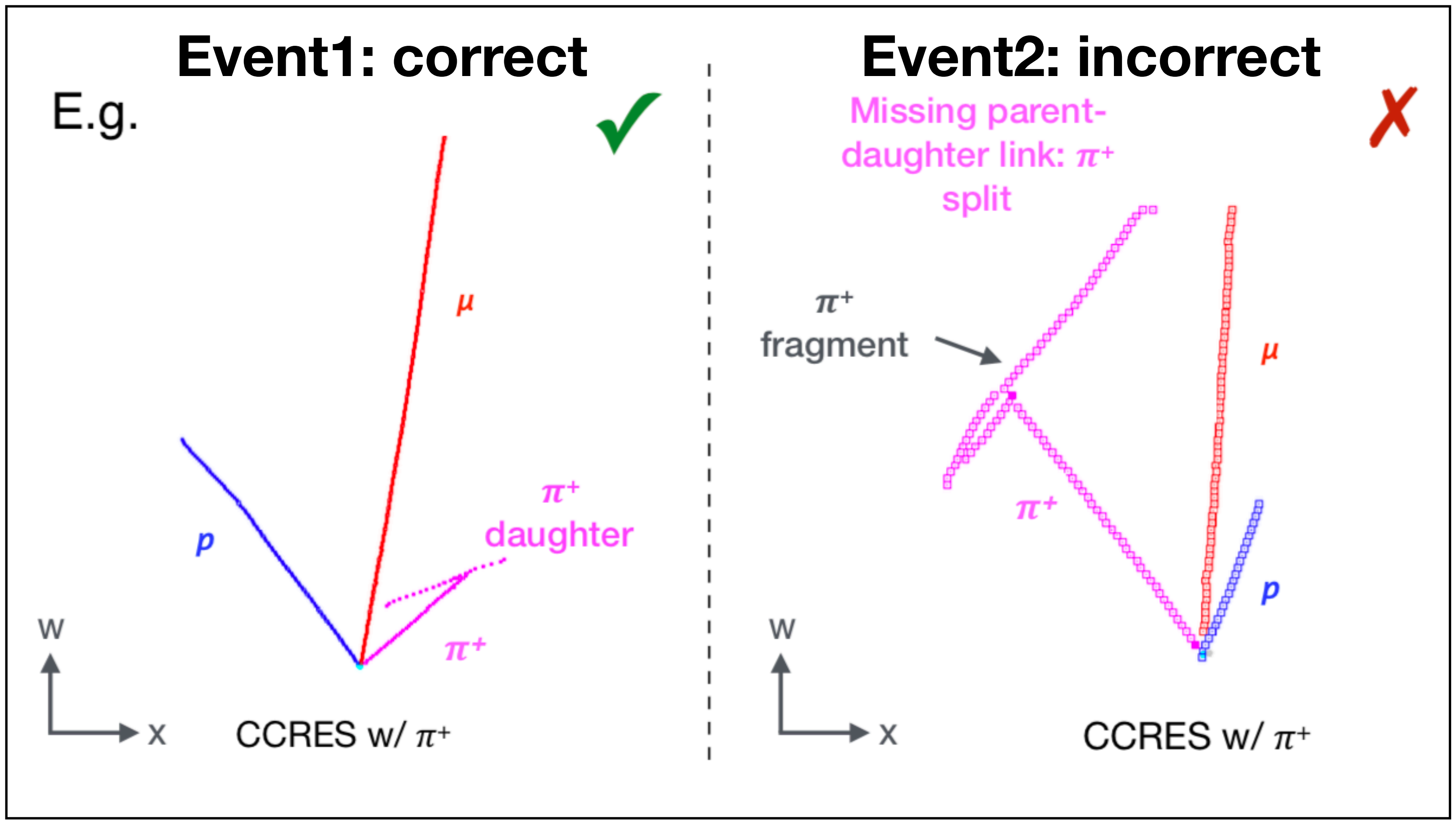
32(+4) PMTs behind the wire planes to record scintillation light and provide trigger information

arxiv:hep-ex/1612.05824 (JINST)

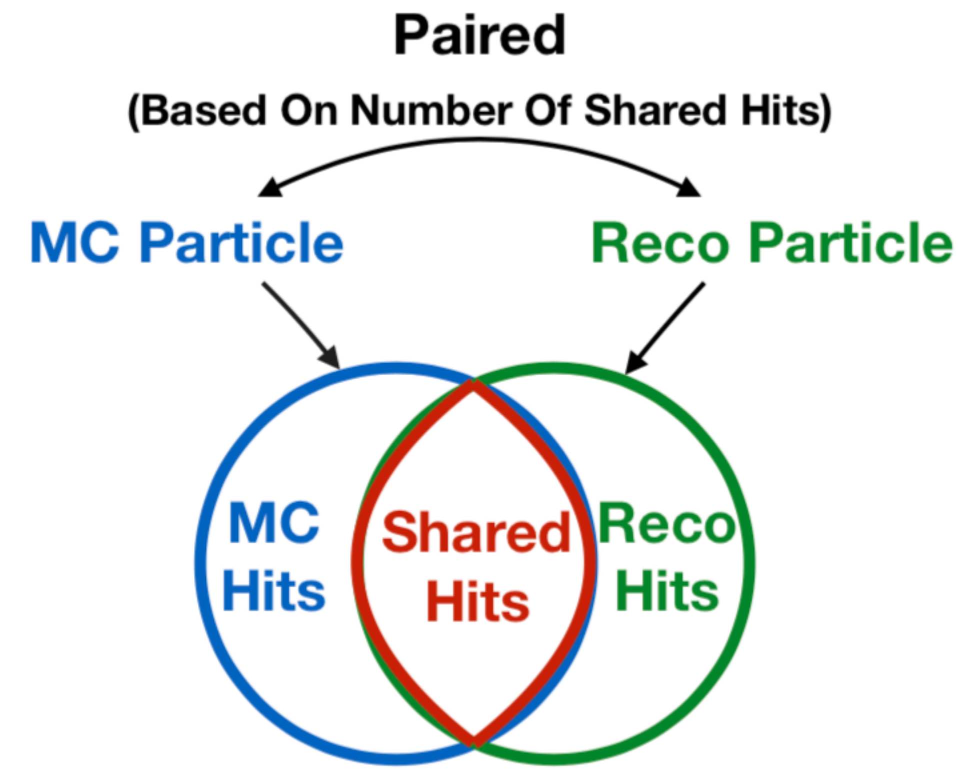
Pandora Performance Metrics

Correct event:
In assessing the performance of the pattern recognition, we only deem an event correct if each target MC Particle (*reconstructable*) is matched to exactly one reconstructed particle. This includes parent-daughter links must be correct, and is a really strict metric.

Efficiency:
Fraction of target MCParticles (*reconstructable) with at least one matched reconstructed particle, where a match needs to fullfill conditions based on number of shared hits:**



***labels correspond to true MC particle type**



*Purity = nSharedHits / nRecoHits > 50%

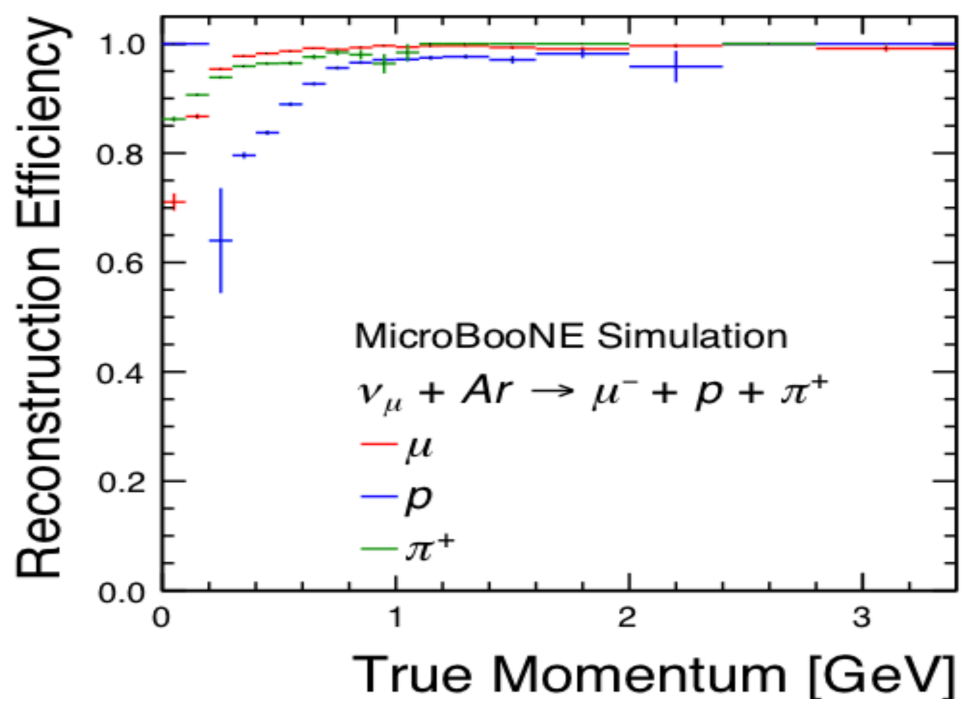
*Completeness = nSharedHits / nMCHits > 10%

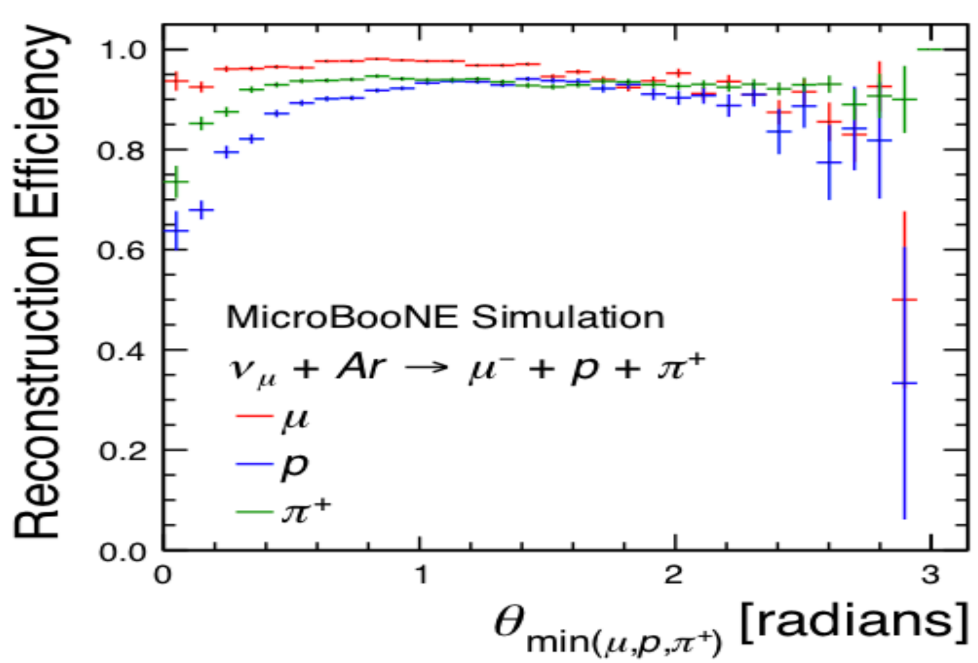
* see more precise definitions in paper

Pandora performance in MicroBooNE

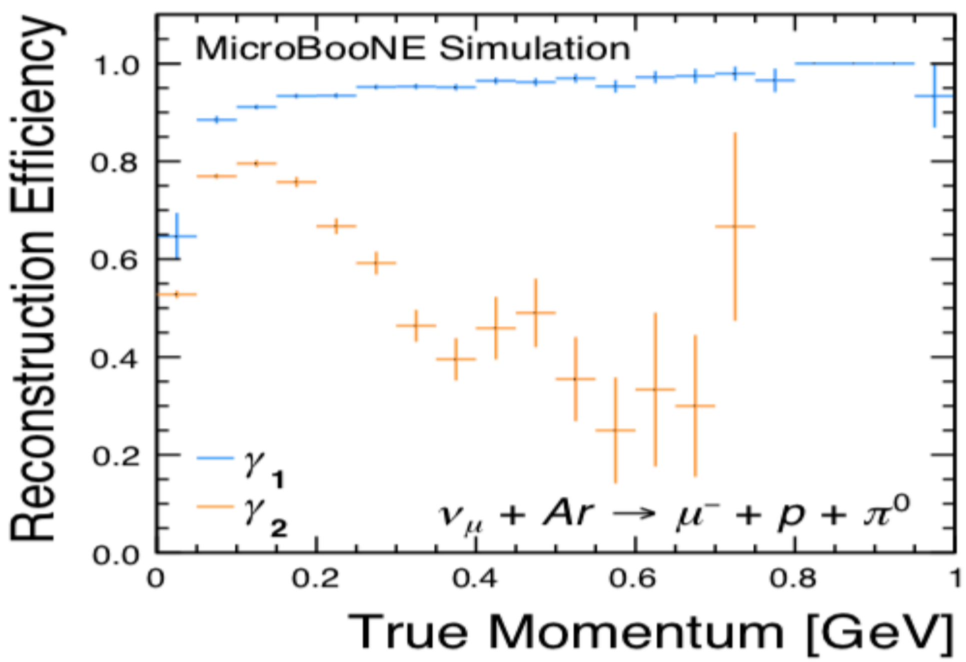
[The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector] **(Eur. Phys. J. C 78, p82 2018)**

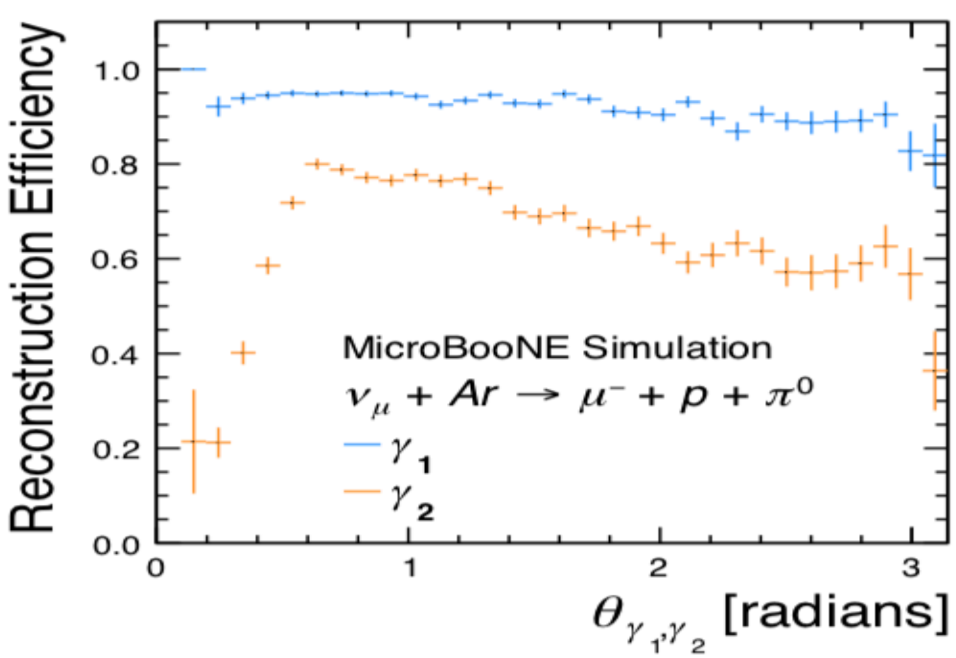
CC RES: μ+p+π^+





CC RES: μ+p+π^0



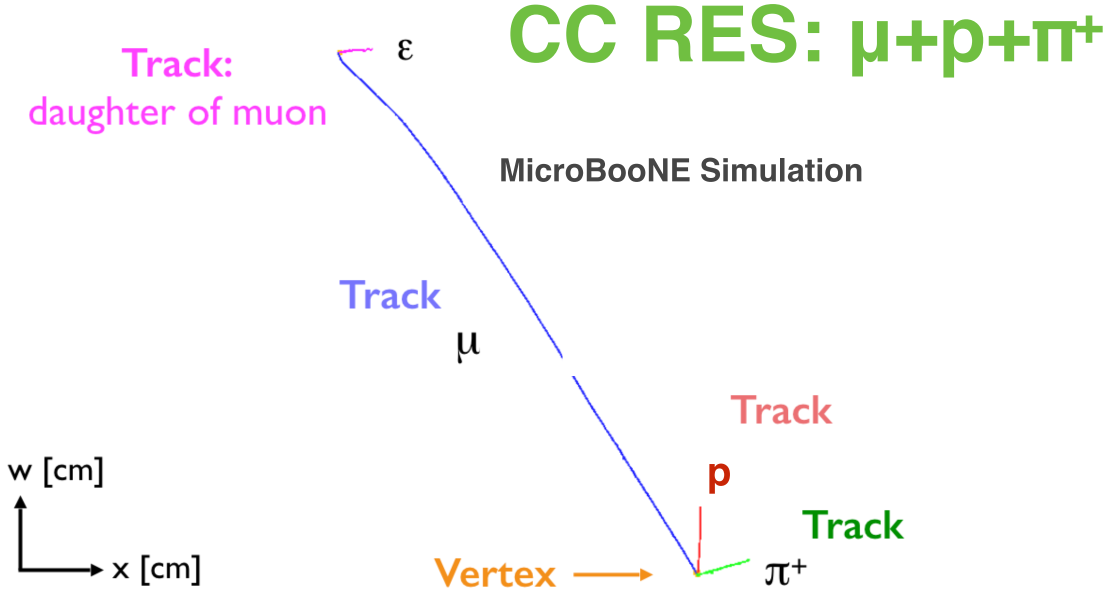


Notice: Pattern recognition stage w/o calorimetry

Pandora performance in MicroBooNE

Examples of matching true-reconstructed particles for different interaction types

CC RES: μ+p+π+



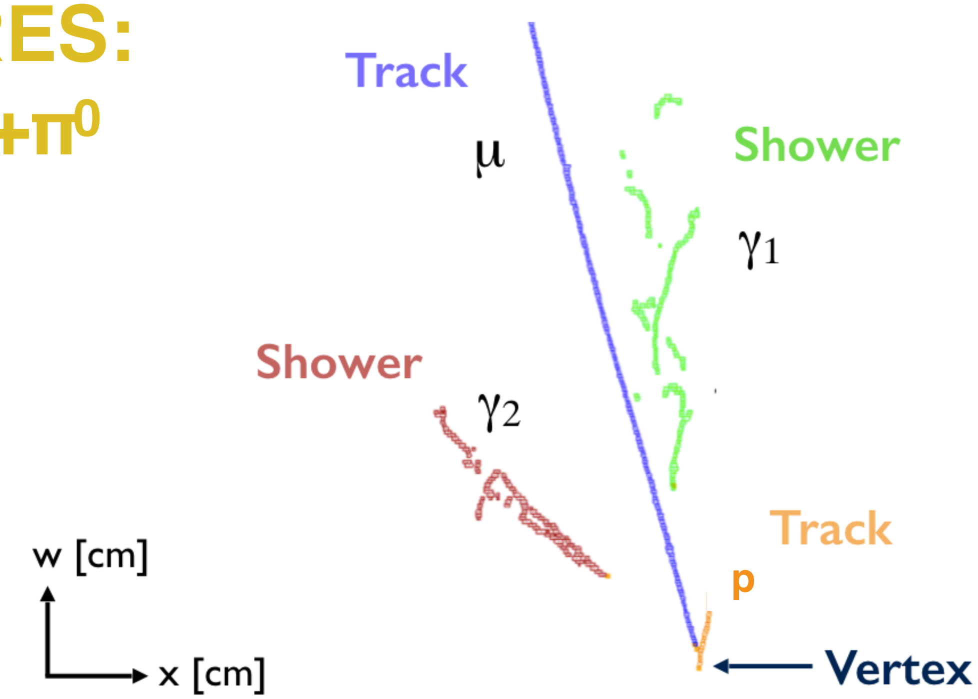
#Matched Particles	0	1	2	3+
μ	3.5%	95.1%	1.4%	0.0%
p	9.0%	86.8%	4.0%	0.3%
π^+	6.9%	80.9%	11.4%	0.8%

0 matches
(not reconstructed or merged with another particle)

=1 match

>1 matches
Particle split

CC RES: μ+p+π0



MicroBooNE Simulatio

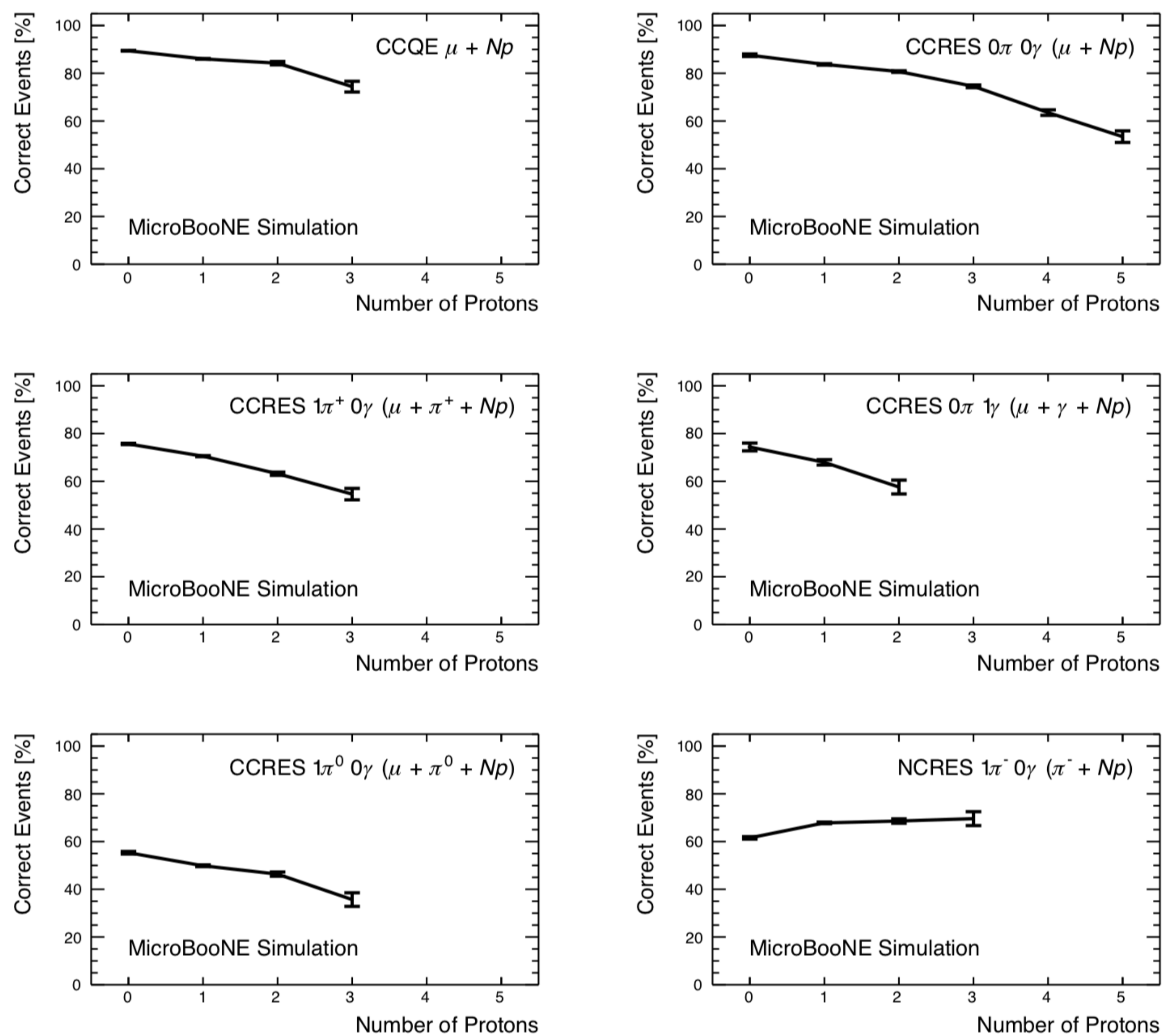
#Matched Particles	0	1	2	3+
μ	3.7%	94.8%	1.5%	0.0%
p	9.9%	85.5%	4.3%	0.3%
γ_1	6.8%	88.0%	4.8%	0.4%
γ_2	29.9%	66.4%	3.6%	0.2%

[The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector]() (Eur. Phys. J. C 78, p82 2018)

Pandora performance in MicroBooNE

The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector **(Eur. Phys. J. C 78, p82 2018)**

Performance on selection of Exclusive Final States



This shows extraordinary possibilities for analyses in different exclusive final state channels in LAr TPCs…

VALOR and oscillation analyses

…which is quite important in oscillation analyses!

"Joint multi-sample analysis is a *necessity*" (Steve Dennis, VALOR)

Steve Dennis

The primary interaction processes that we want to study and constrain are **mingled** together due to the presence of the nucleus.

Final-State	Primary Hadronic System									
	$0\pi X$	$1\pi^0 X$	$1\pi^+ X$	$1\pi^- X$	$2\pi^0 X$	$2\pi^+ X$	$2\pi^- X$	$\pi^0\pi^+ X$	$\pi^0\pi^- X$	$\pi^+\pi^- X$
$0\pi X$	**293446**	12710	22033	3038	113	51	5	350	57	193
$1\pi^0 X$	1744	**44643**	3836	491	1002	25	1	1622	307	59
$1\pi^+ X$	2590	1065	**82459**	23	14	660	0	1746	5	997
$1\pi^- X$	298	1127	1	**12090**	16	0	46	34	318	1001
$2\pi^0 X$	0	0	0	0	**2761**	2	0	260	40	7
$2\pi^+ X$	57	5	411	0	1	**1999**	0	136	0	12
$2\pi^- X$	0	0	0	1	0	0	**134**	0	31	0
$\pi^0\pi^+ X$	412	869	1128	232	109	106	0	**9837**	15	183
$\pi^0\pi^- X$	0	0	1	0	73	0	8	5	**1808**	154
$\pi^+\pi^- X$	799	7	10	65	0	0	0	139	20	**5643**

Left: Migration from "primary" (before FSI) to final-state / observed topologies ($\nu_\mu O^{16}$, 1 GeV)

*** Raw GENIE (produced by Costas)**

For example, the 0π, $1\pi^+$, $\pi^+\pi^0$, $2\pi^+$ and other datasets are connected via the *same pion FSI physics*: In a joint multi-channel analysis, one would not be allowed to pull FSI parameters to "fix" an unrelated problem manifesting in one of the channels.

VALOR: Neutrino fitting group doing oscillation analysis in different experiments, T2K, DUNE, SBN (https://valor.pp.rl.ac.uk)

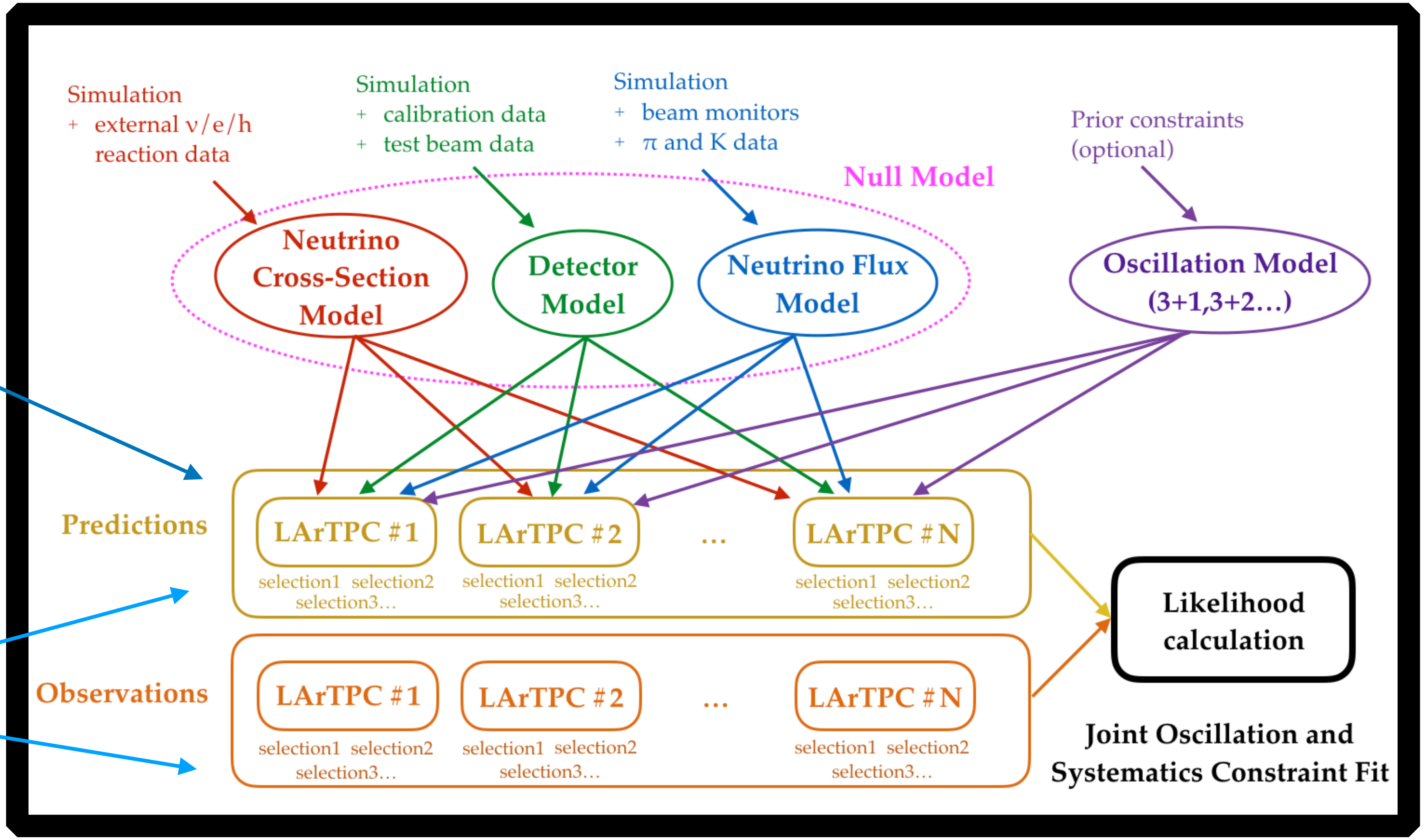
VALOR and oscillation analyses

See poster at Neutrino2016 here

VALOR Neutrino Fitting Group Strategy for a Joint Oscillations Fit of Multiple LArTPCs

Multiple detectors
(“If it goes in the pit, it can go in the fit”, Steve Dennis)

Multiple samples
(broken down into different final state channels)



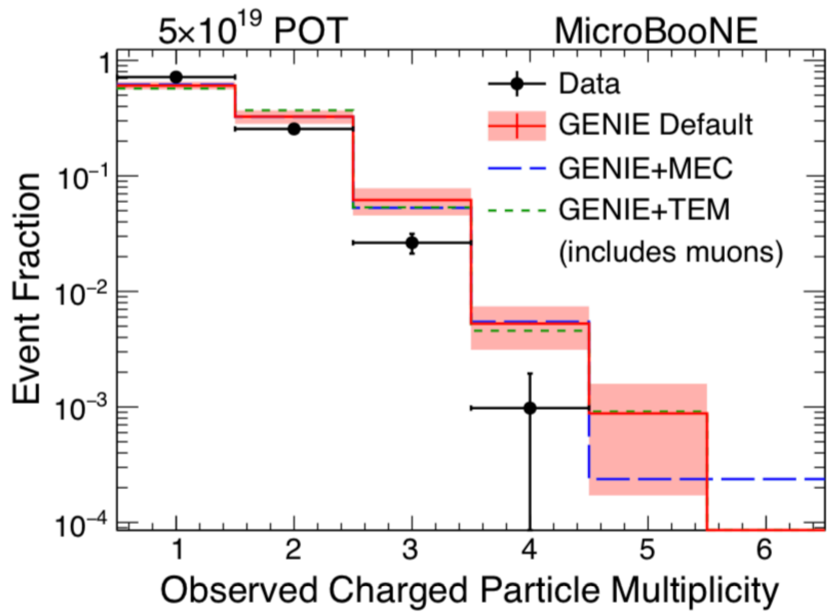
Pandora & MicroBooNE Analyses

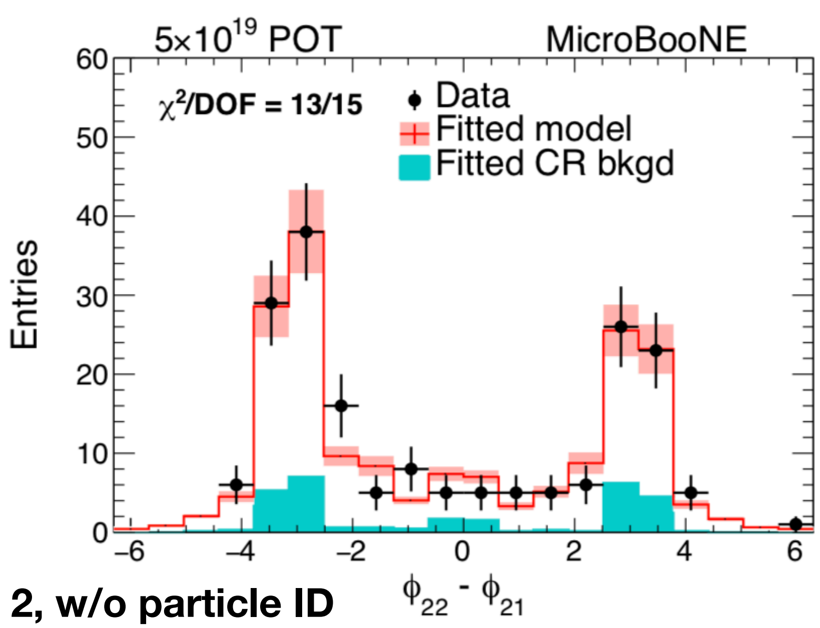
Pandora is used in multiple analyses in MicroBooNE, I have just selected a few representative results (see more in MicroBooNE public docs web)

INCLUSIVE

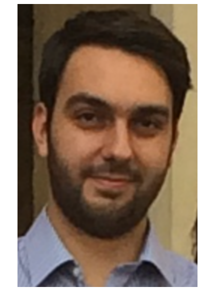
Aleena Rafique

Comparison of ν_μ –Ar multiplicity distributions observed by MicroBooNE to GENIE model predictions (paper)



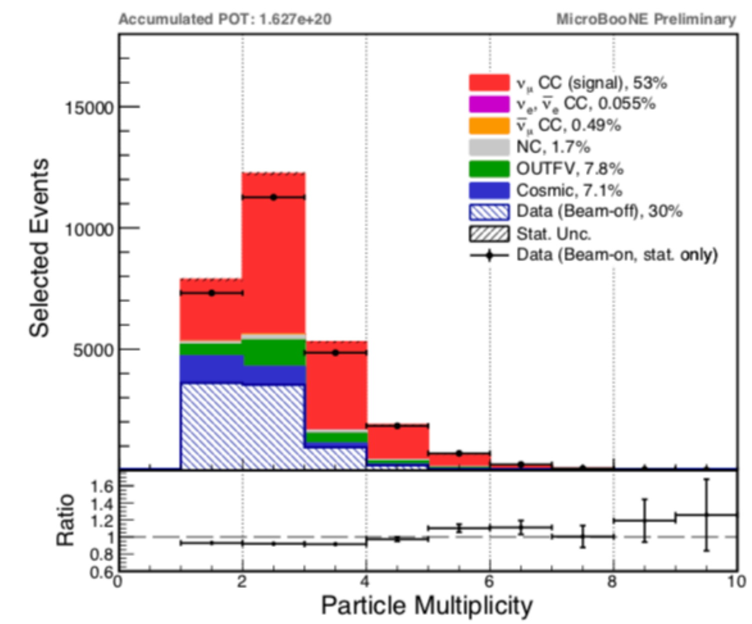


Multiplicity 2, w/o particle ID

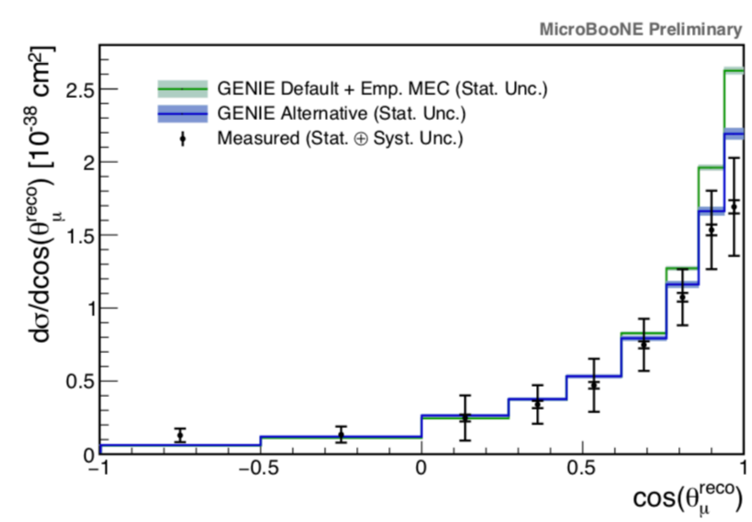


Marco del Tutto

First ν_μ CC Inclusive Differential Cross Section Measurement for MicroBooNE Run 1 Data (public note)



#Reconstructed particles from neutrino vertex in selected events



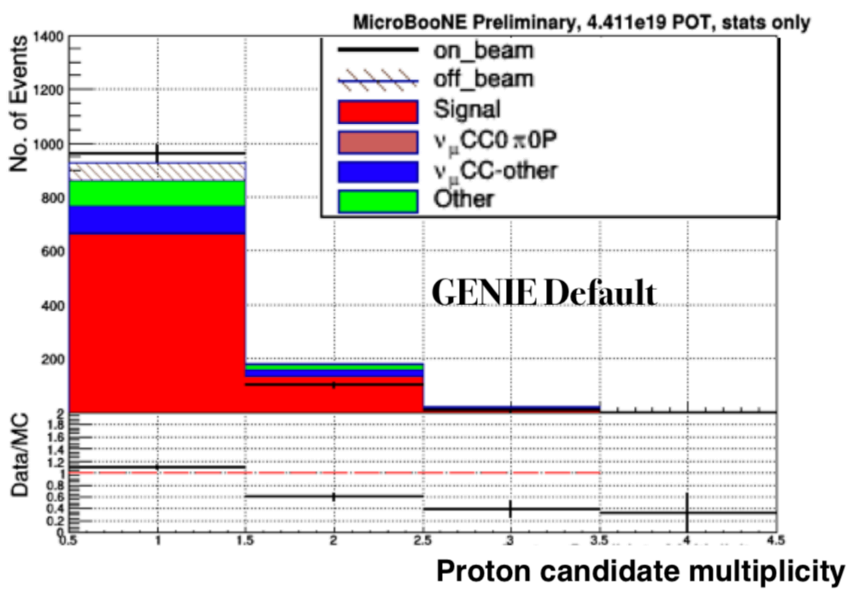
Pandora & MicroBooNE Analyses

Pandora is used in multiple analyses in MicroBooNE, I have just selected a few representative results (see more in MicroBooNE public docs [web](web))

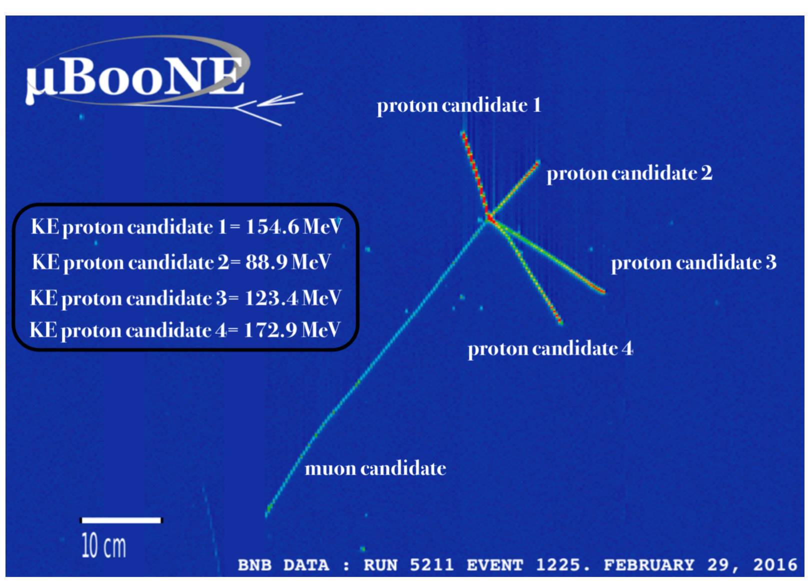
EXCLUSIVE

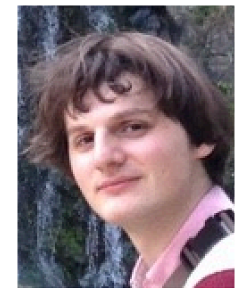
Raquel Castillo

MicroBooNE Charged-Current Analyses with final state protons (NuInt talk)



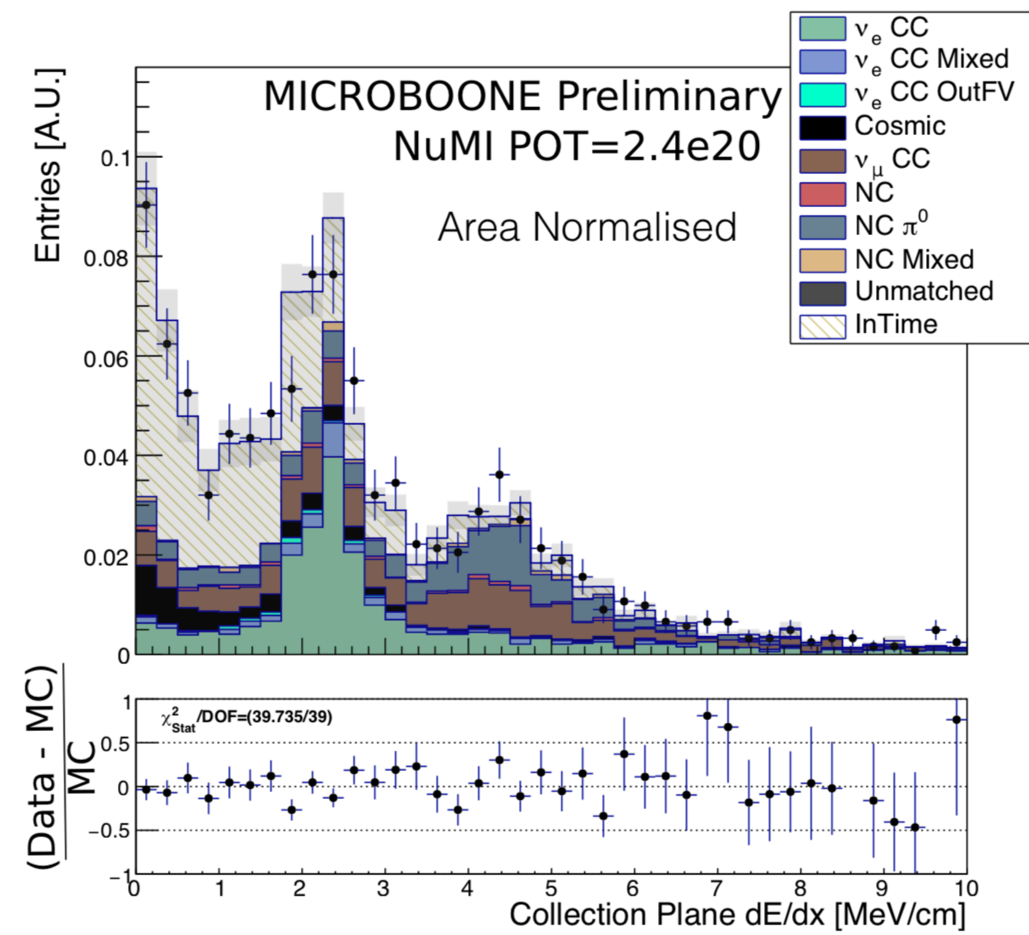
Proton multiplicity in ν_μ CC 0π selected events





Colton Hill

Automated Selection of ve from NuMI using MicrooBooE (NuInt talk)



Pandora & Detector Physics

Pandora reconstruction is also used in calibration and detector physics studies, for example:

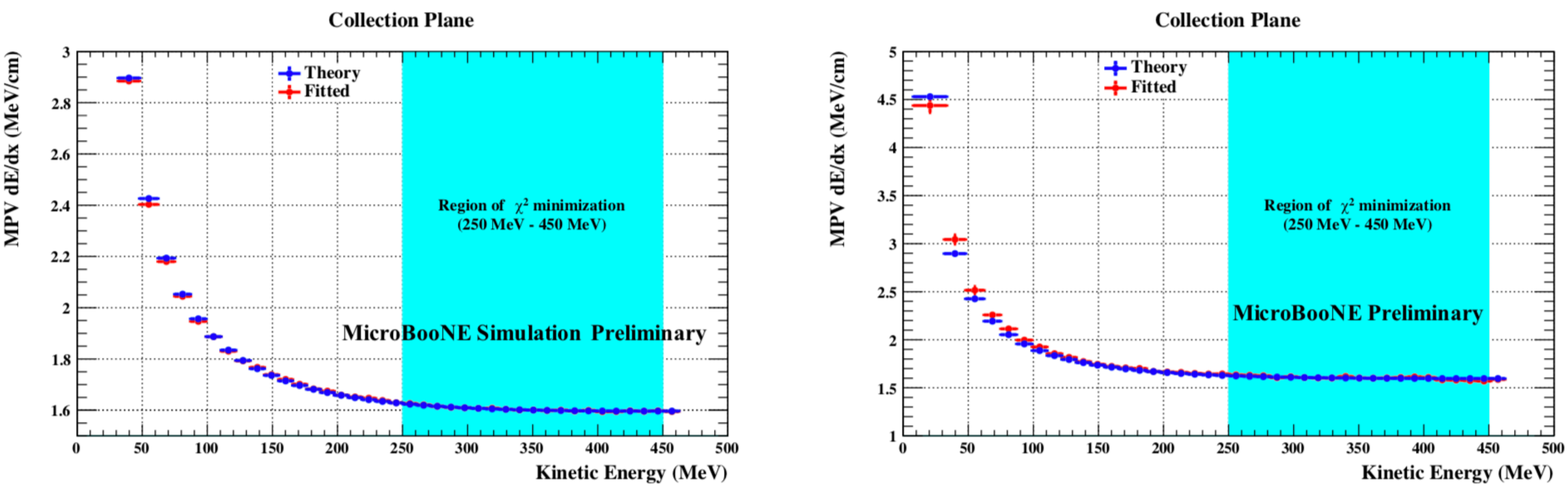
Detector calibration using through going and stopping muons in the MicroBooNE LArTPC (public note)

Varuna Meddage

David Caratelli

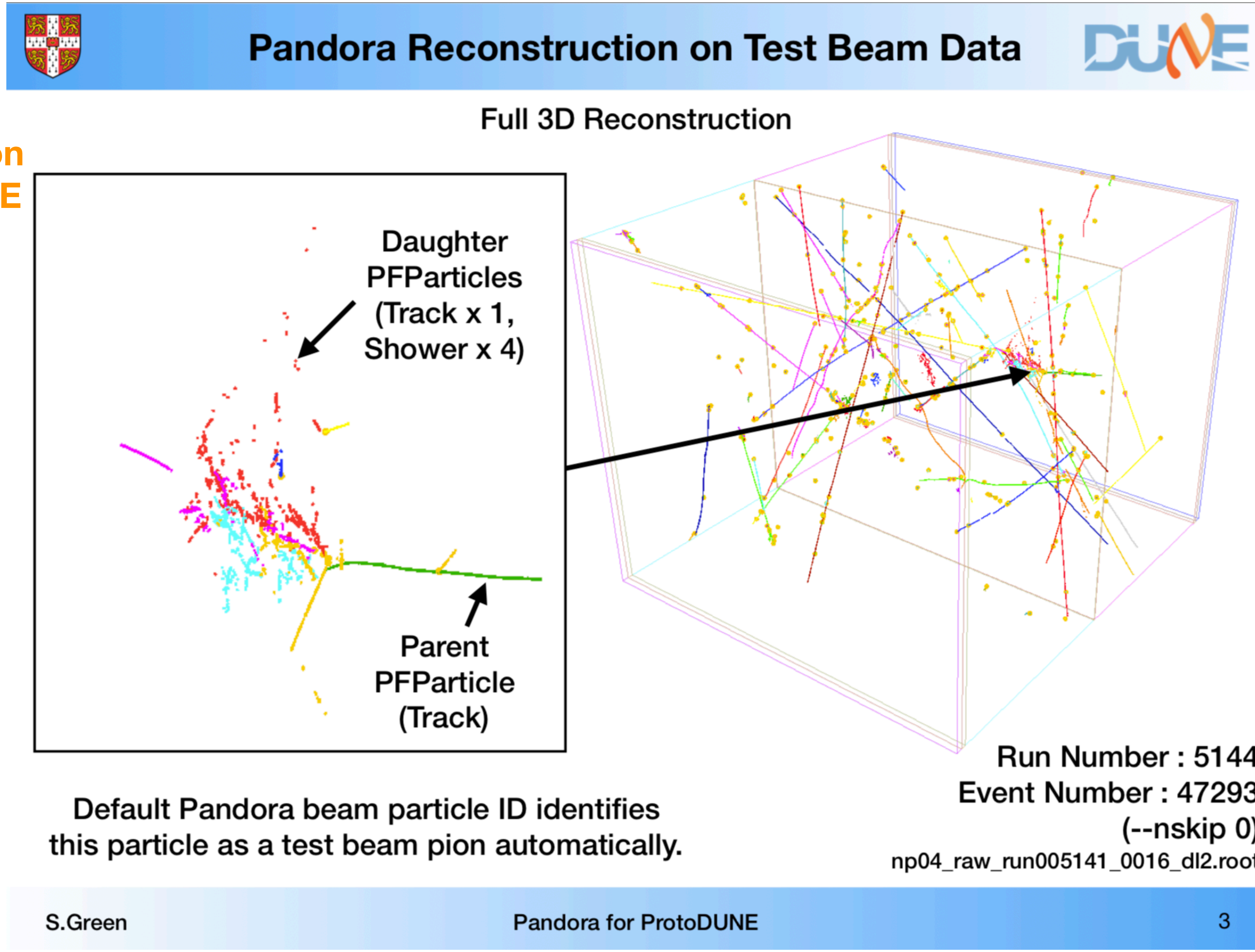
Tingjun Yang

+ others!



Predicted and fitted most provable value of dEdx in the collection plane for stopping muons in MC (left) and data (right)

Exciting times for protoDUNE!

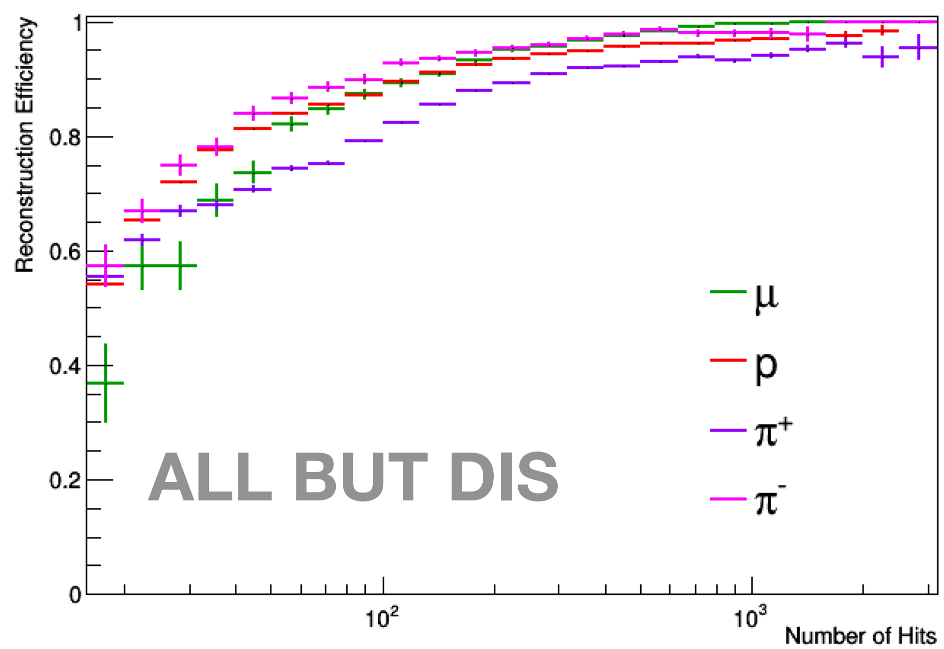


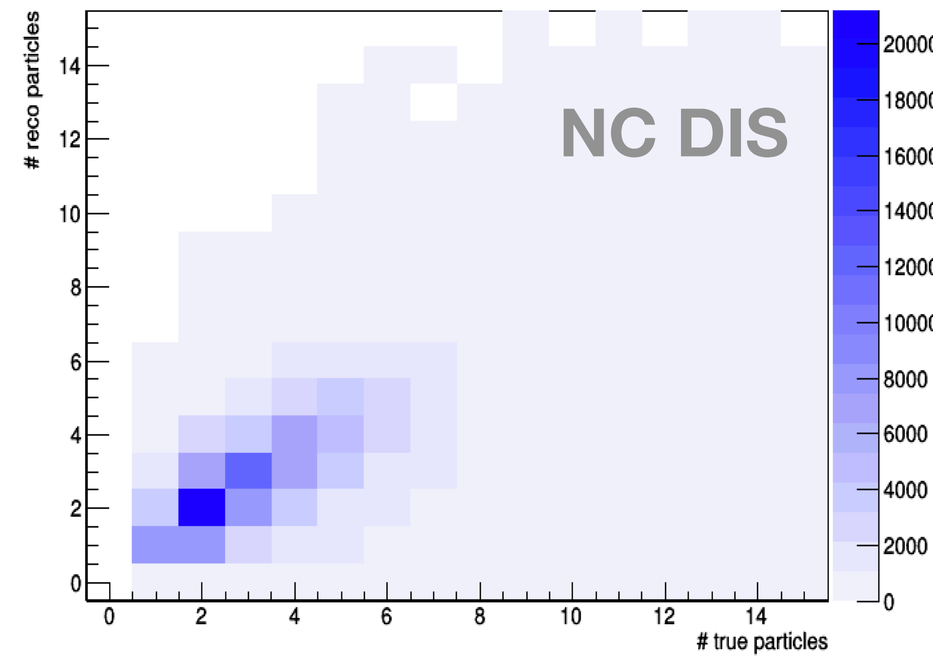
Pandora reconstruction on protoDUNE real data!

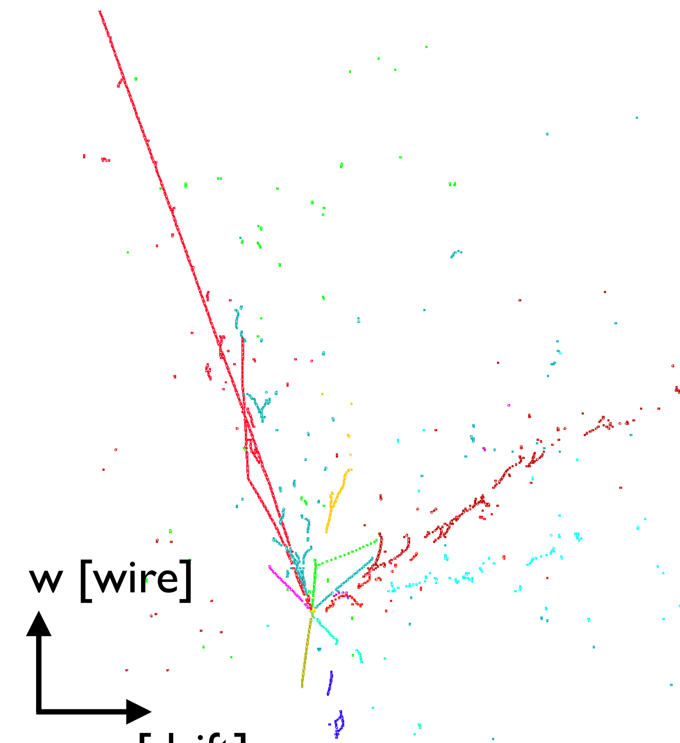
Steven Green

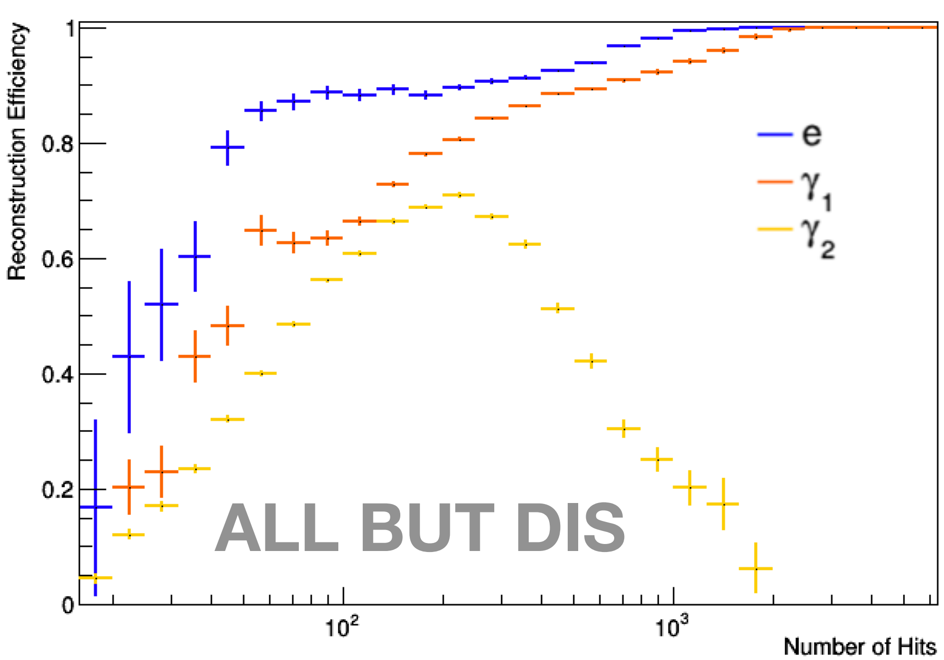
And getting ready for DUNE FD

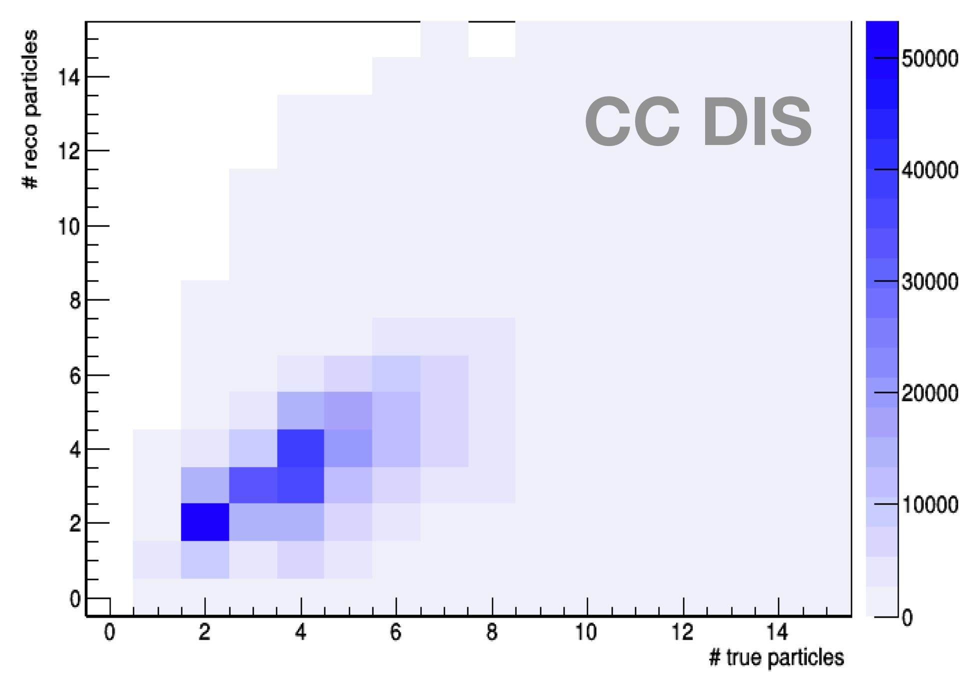
Preliminary performance on DUNE FD (1x2x6) simulation











Lorena Escudero

Note: Total number of hits (U+V+W)

So stay tuned!

Summary & Plans

- The Pandora multi-algorithm approach enables us to build up reconstructed images of very complex interactions
- Pandora uses both pure 2D and pure 3D approaches, in addition to algorithms where multiple 2D points are used to find test 3D positions, or 3D positions are projected into 2D
- Machine Learning methods can be incorporated in this approach to drive algorithm decisions.
- Full particle hierarchies are delivered, with tagging of cosmic-ray muons and neutrinos/test beam particles providing a consolidated output
- Pandora delivers necessary output for physics analysis that really exploit imaging capabilities of LArTPCs!

We are actively working on the SBN program (mainly MicroBooNE) and protoDUNE and DUNE, and expecting to grow the team in 2019!

Thanks!

Pandora is an open project and new contributors would be extremely welcome.
We'd love to hear from you and we will always try to answer your questions.

Pandora SDK Development	John Marshall (John.Marshall@warwick.ac.uk) Mark Thomson (thomson@hep.phy.cam.ac.uk)
LAr TPC algorithm development	John Marshall (John.Marshall@warwick.ac.uk) Andy Blake (a.blake@lancaster.ac.uk)
DUNE FD Integration	Lorena Escudero (escudero@hep.phy.cam.ac.uk)
ProtoDUNE Integration	Steven Green (sg568@hep.phy.cam.ac.uk)
MicroBooNE Integration	Andy Smith (asmith@hep.phy.cam.ac.uk)
Other team members	MicroBooNE: Joris Jan de Vries, Jack Anthony ProtoDUNE: Stefano Vergani

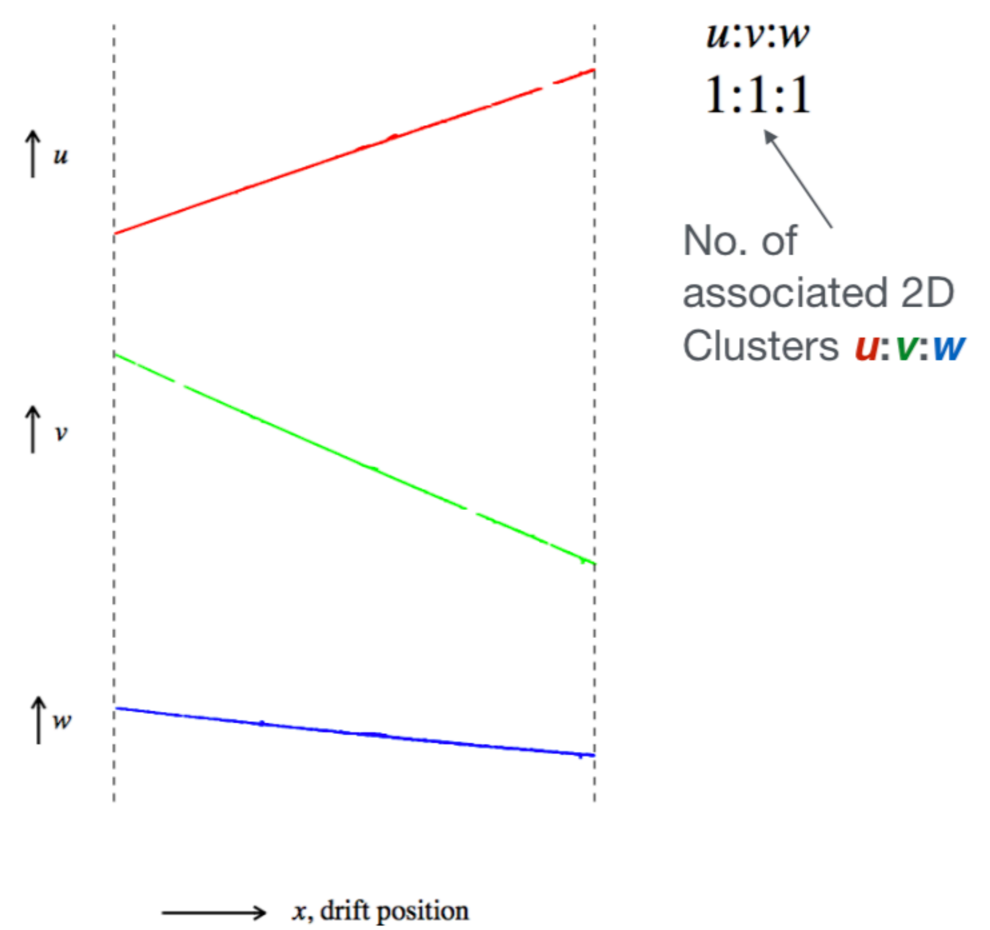
https://github.com/PandoraPFA

https://pandorapfa.slack.com

Pandora LAr TPC Pattern Recognition

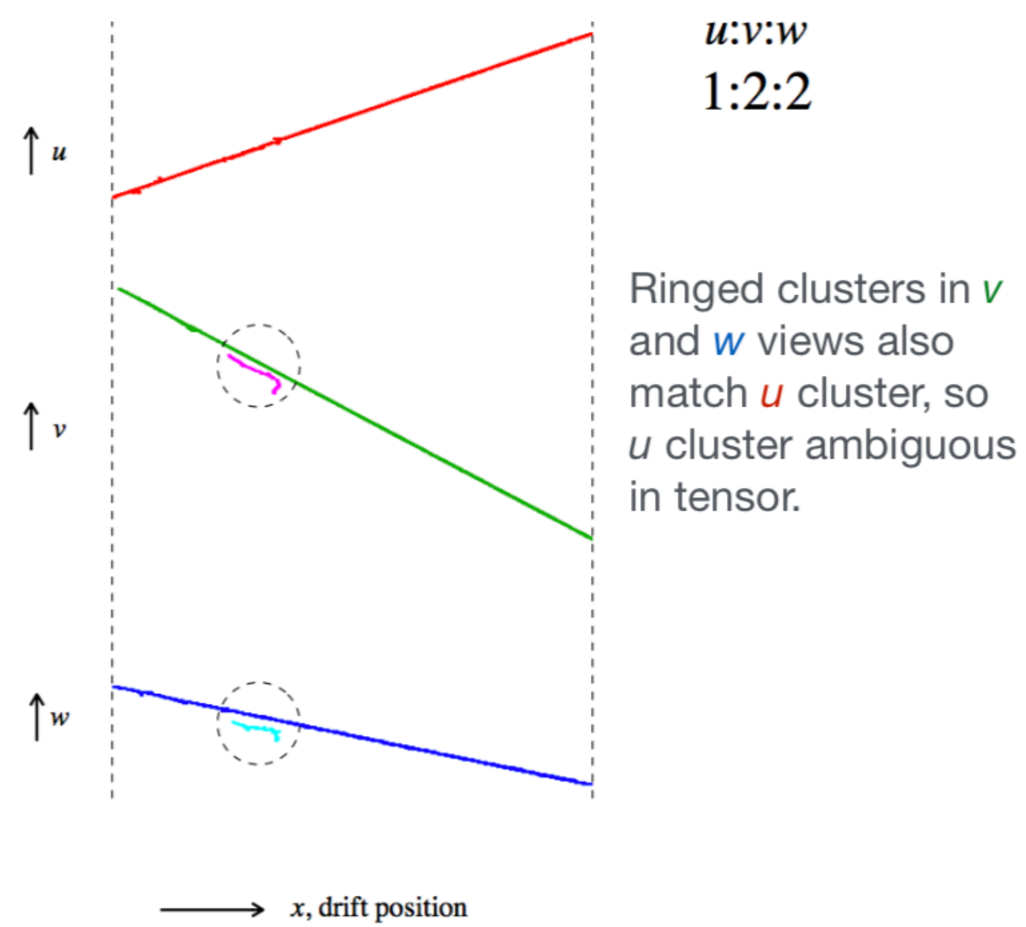
Tensor ambiguities example

- Tensor stores overlap details for trios of 2D clusters. Tools make 2D reco changes to resolve any ambiguities. If a tool makes a change (e.g. splits a cluster), all tools run again.



Find unambiguous elements in the tensor, demanding that the common x-overlap is 90% of the x-span for all three clusters.

Clear Tracks Tool



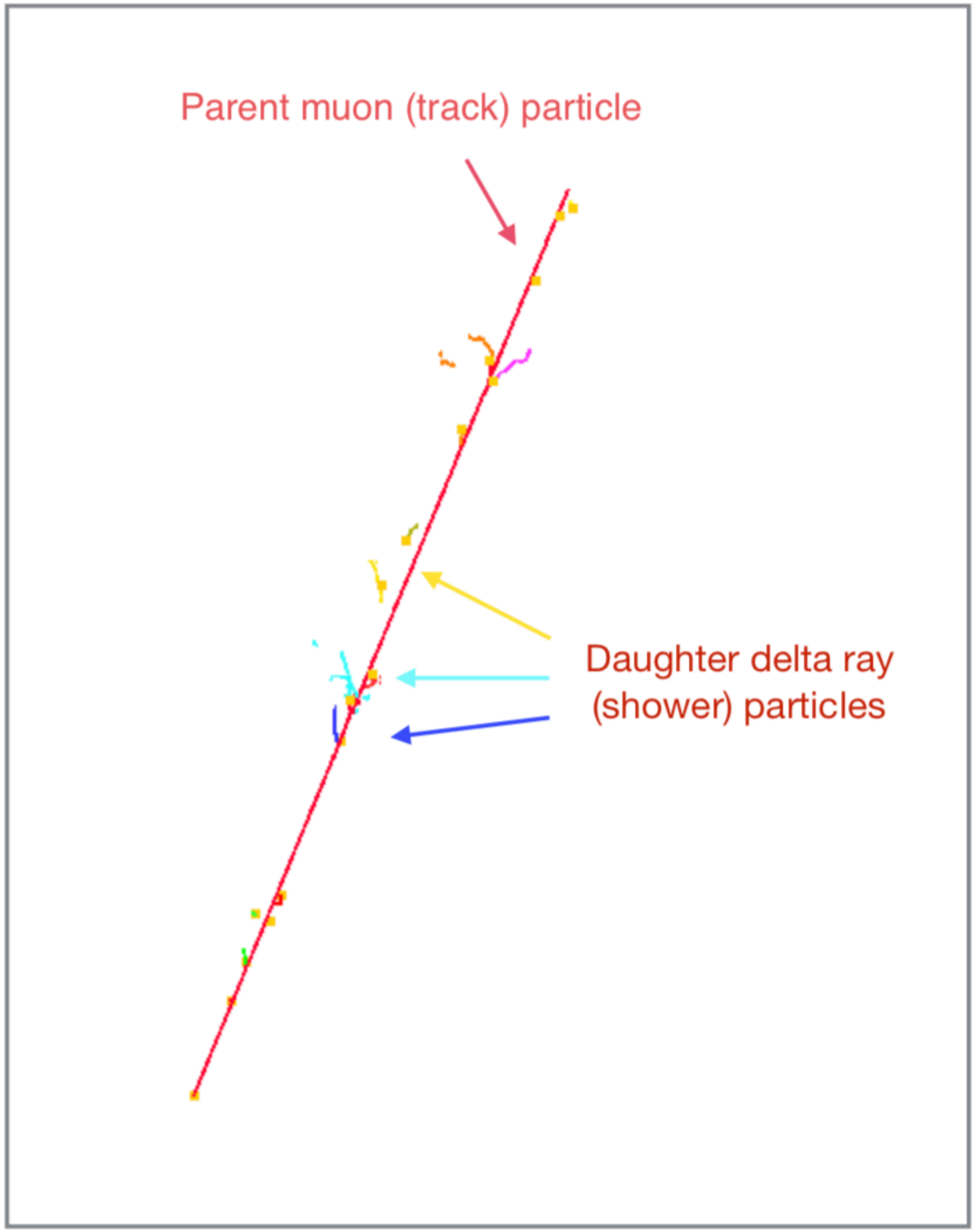
Resolve **obvious** ambiguities: clusters are matched in multiple configurations, but one tensor element is much "better" than others.

Long Tracks Tool

Pandora LAr TPC Pattern Recognition

Delta-ray Reconstruction

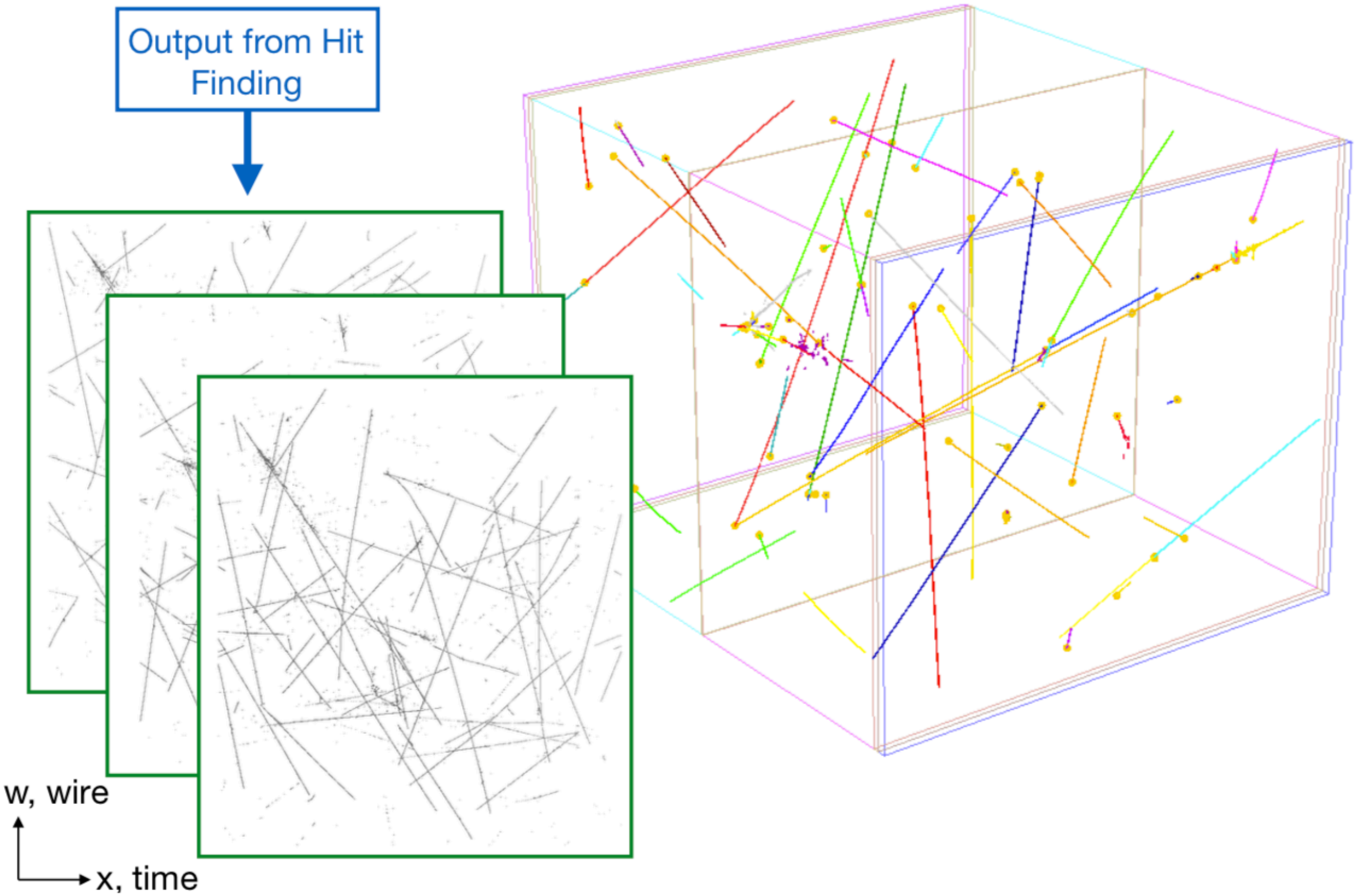
- Assume any 2D clusters not in a track particle are from delta-ray showers:
  - Simple proximity-based reclustering of hits, then topological association algs.
  - Delta-ray clusters matched between views, creating delta-ray shower particles.
  - Parent muon particles identified and delta-ray particles added as daughters.

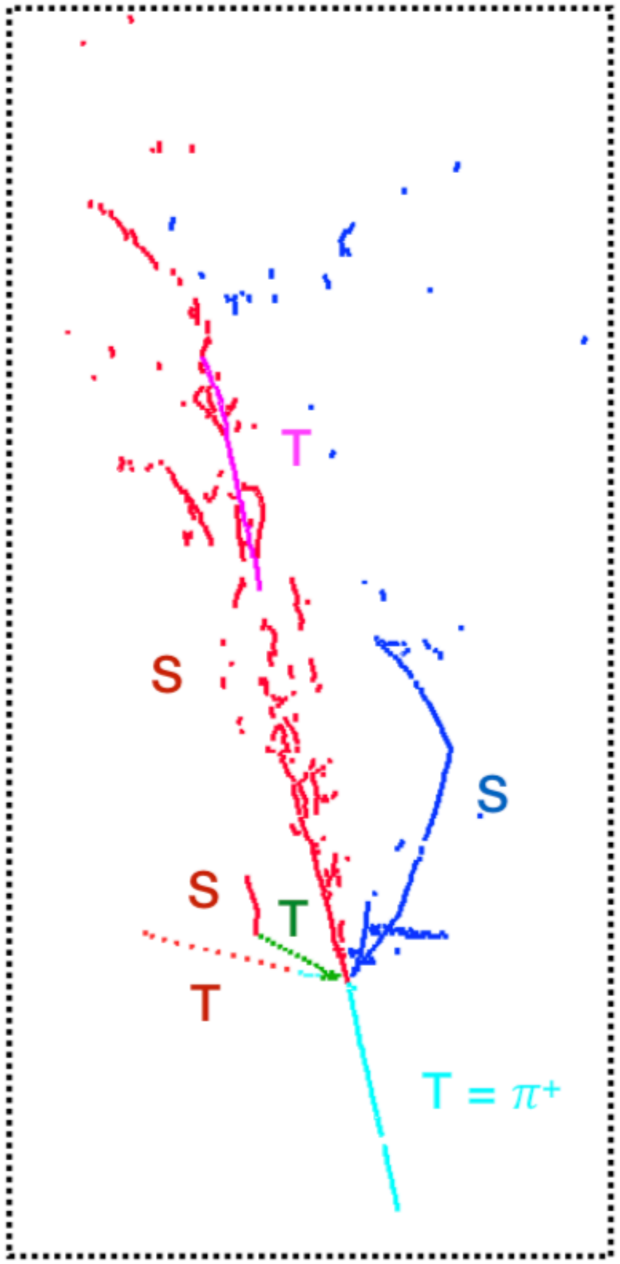


Pandora LAr TPC Pattern Recognition

Aim of the pattern recognition:

- **Take the reconstructed hits (previous stage)**
- **Reconstruct 3D particles from 3x2D views**
- **Produce hierarchy of particles in interaction**





Simulated π^+ Pandora Reconstruction at ProtoDUNE-SP

Pandora LAr TPC Pattern Recognition

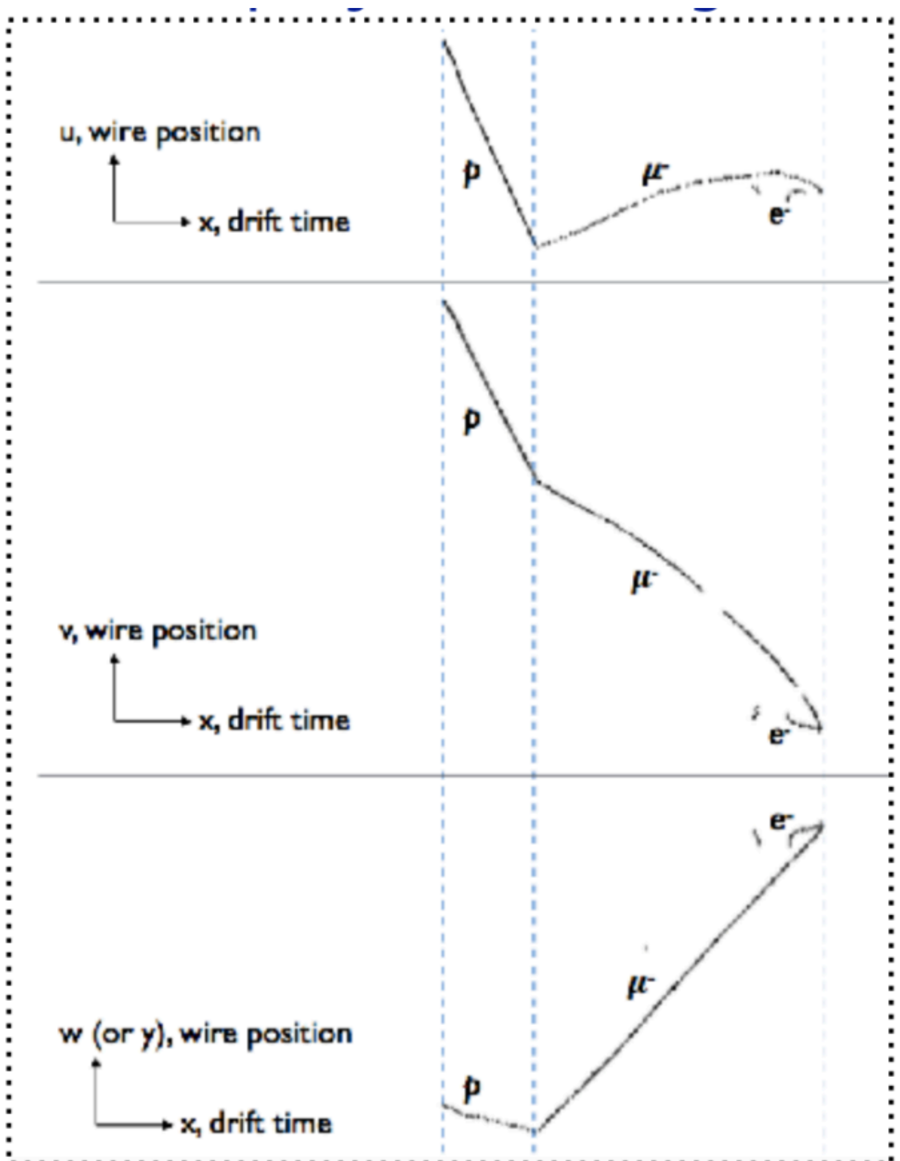
How to do that? A one slide summary

Pattern recognition problem in 3 steps:

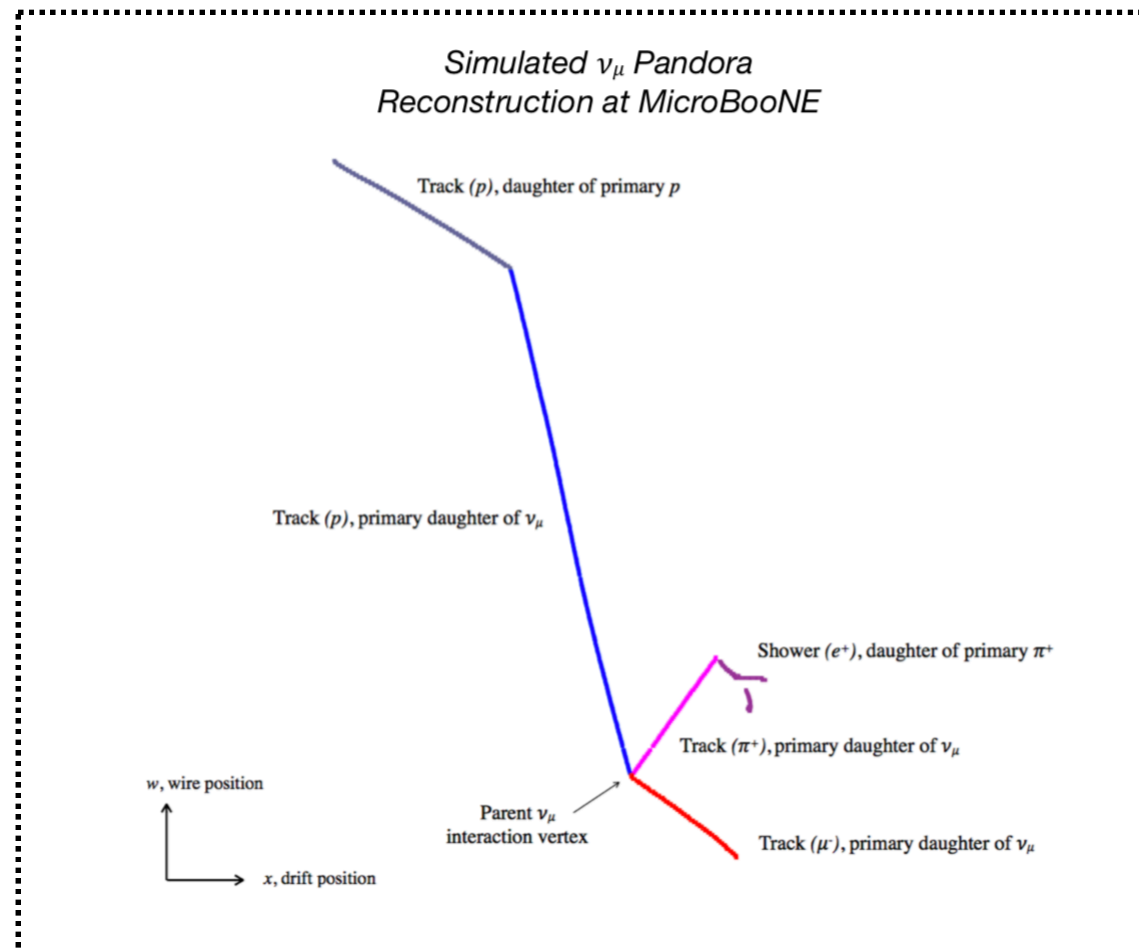
1) Identify features and cluster hits in 2D images

2) Reconstruct 3D particles from 2D images

3) Reconstruct particle hierarchies or "flow" (parent-daughters)



This is our input: a collection of hits in the 2D views (and detector information such as dead channels if applicable)

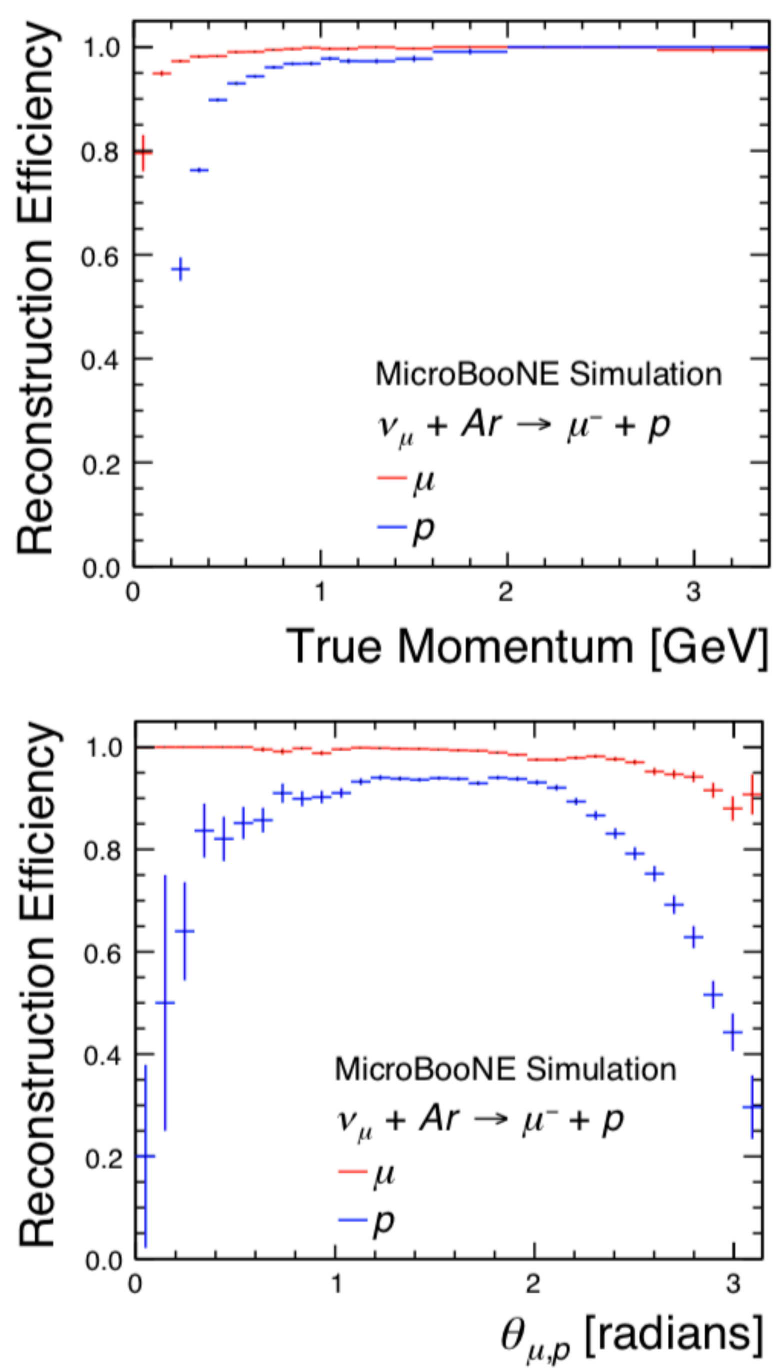


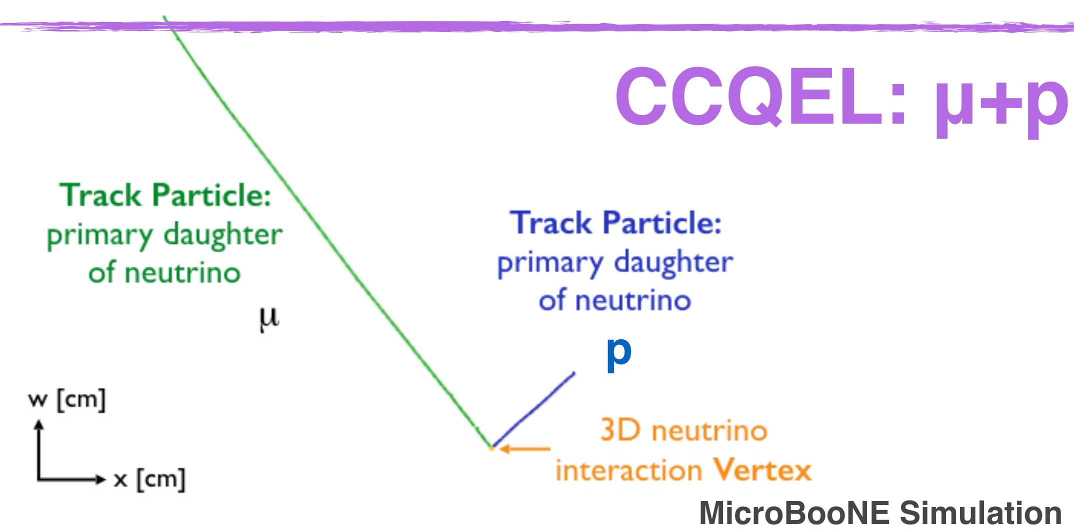
This is our output: the hierarchy of particles in the interaction occurred

Pandora performance in MicroBooNE

[The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector]() **(Eur. Phys. J. C 78, p82 2018)**

CCQEL: μ+p





#Matched Particles	0	1	2	3+
μ	1.3%	95.8%	2.9%	0.1%
p	8.9%	87.3%	3.6%	0.2%

Notice: Pattern recognition stage w/o calorimetry

Pandora LAr TPC Pattern Recognition

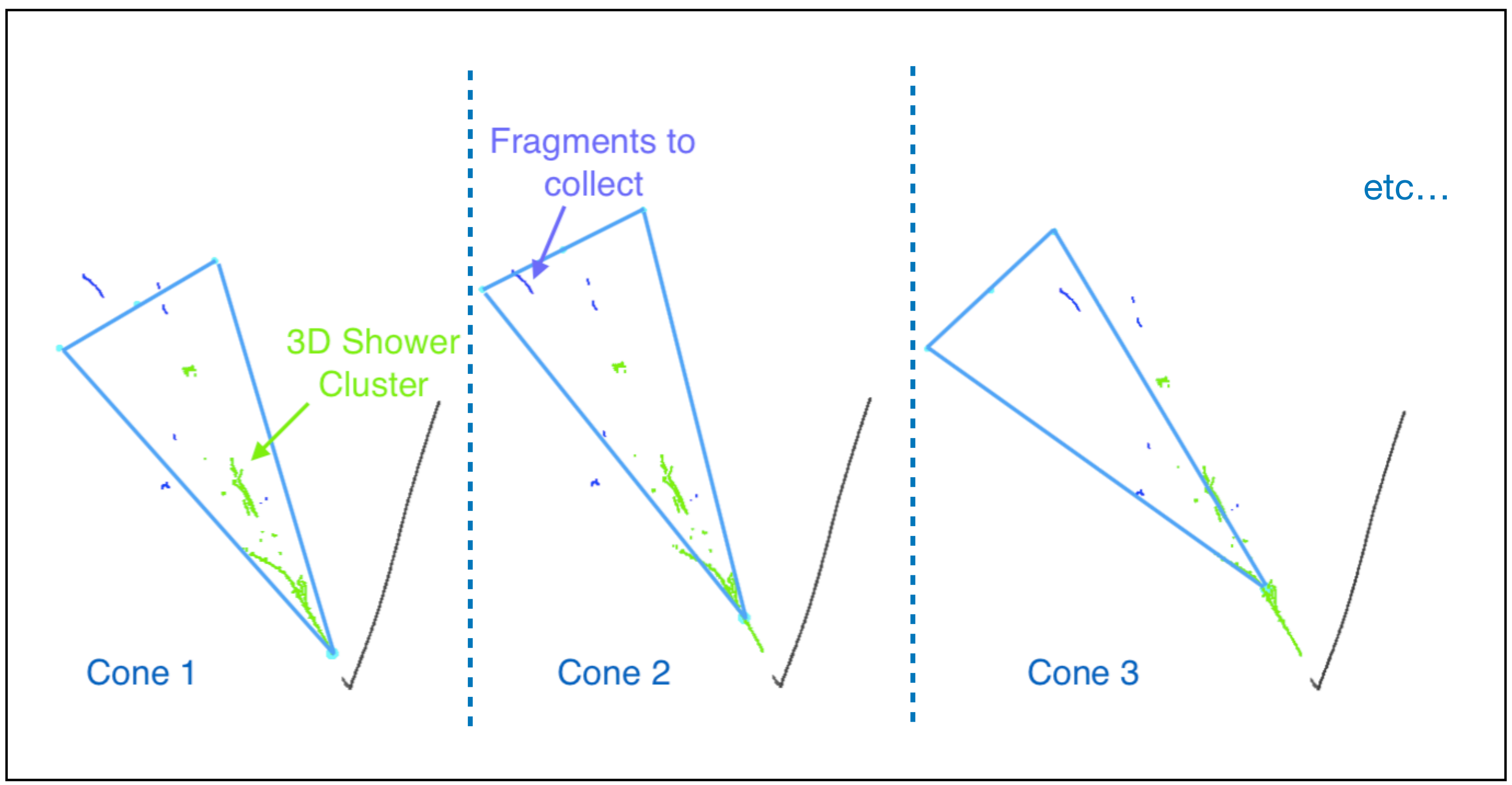
A little bit deeper into some of the steps in the algorithm chains…

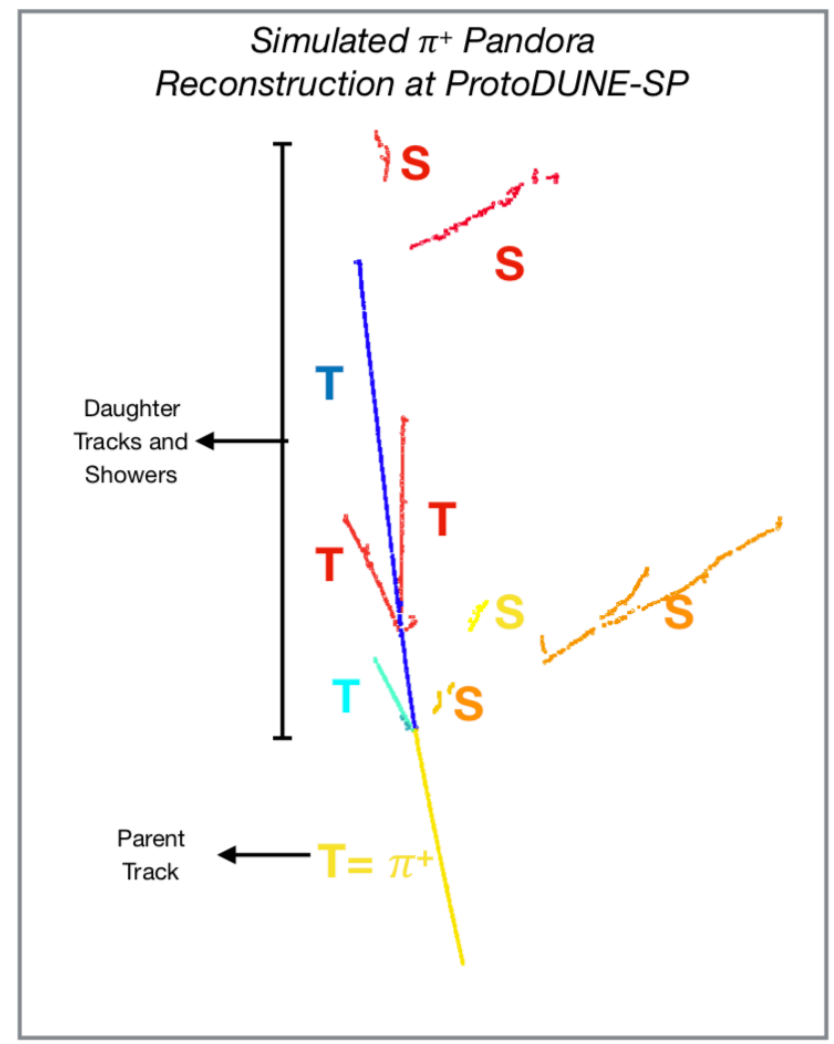
2D/3D Particle refinement

Several algorithms deal with remnants to improve particle completeness, (esp. sparse showers). Sliding linear fits are used to define 2D envelopes and 3D cones for picking up small clusters/fragments.

and 3D particle hierarchy

Finally, walking backwards from interaction vertex, use 3D clusters to organise particles into hierarchies (building parent-daughter links)





Can't do justice in a few slides, please find more in the Pandora MicroBooNE paper (Eur. Phys. J. C 78, p82 2018)